

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 18, 2002, 11:30:36 ; Search time 35 seconds
(without alignments)
2278.292 Million cell updates/sec

Title: US-09-895-913a-120
Perfect score: 387
Sequence: 1 MIPKMERALGSGVILSKD.....KPKRFLVLDLNOGYRIILVK 387

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	387	100.0	443	16	Q25663 helicobacte
2	96	24.8	476	16	Q92M18 helicobacte
3	20	5.2	368	2	Q46094 campylobact
4	20	5.2	472	2	Q46120 campylobact
5	20	5.2	472	16	Q9PN69 campylobact
6	18	4.7	371	2	Q31388 bradyrhizob
7	18	4.7	481	16	Q9PGL3 xylella fas
8	17	4.4	495	16	Q8UE46 agrobacteri
9	14	3.6	348	16	Q97LU1 clostridium
10	14	3.6	367	16	Q8RGV1 thermococ
11	13	3.4	519	16	Q93J30 streptomyc
12	12	3.1	424	16	Q99TD6 staphylococ
13	12	3.1	437	2	Q68197 haemophilus
14	12	3.1	464	2	Q56885 yersinia en
15	12	3.1	474	2	Q44596 brucella ab
16	12	3.1	474	16	Q8YHL4 brucella me

17	12	3.1	475	16	Q8Z9B0
18	12	3.1	477	2	Q8RSS1
19	12	3.1	478	2	P74978
20	12	3.1	481	16	Q8ZBM6
21	11	2.8	347	16	Q9KZU4
22	11	2.8	428	16	Q98N31
23	11	2.8	447	16	Q8R756
24	11	2.8	453	2	Q9FD11
25	11	2.8	455	16	Q8ZLQ1
26	11	2.8	455	16	Q8Z3E6
27	11	2.8	455	16	Q8X9F1
28	11	2.8	457	16	Q8ZB58
29	11	2.8	460	2	O06439
30	11	2.8	491	2	Q9LBK0
31	11	2.8	499	16	Q9ZEX8
32	11	2.8	500	16	Q8YA67
33	10	2.6	342	16	Q9CMR4
34	10	2.6	362	16	Q9S2K5
35	10	2.6	382	16	Q9CD67
36	10	2.6	389	2	Q51374
37	10	2.6	389	16	Q9HVX1
38	10	2.6	403	16	Q8XV99
39	10	2.6	406	16	Q9KSR6
40	10	2.6	407	16	Q99XG9
41	10	2.6	446	16	Q8VKA4
42	10	2.6	452	2	Q9Z5G6
43	10	2.6	456	16	Q9KUF5
44	10	2.6	464	16	O53896
45	10	2.6	468	16	Q9Z282

ALIGNMENTS

RESULT 1

Q25663	ID	Q25663	PRELIMINARY;	PRT;	443 AA.
AC	O25663;				
DC	01-JAN-1998	(TREMREL. 05, Created)			
DT	01-JAN-1998	(TREMREL. 05, Last sequence update)			
DT	01-MAR-2002	(TREMREL. 20, Last annotation update)			
DE	Serine protease (HTRA).				
GN	HPI019.				
OS	Helicobacter pylori (Campylobacter pylori).				
OC	Bacteria; proteobacteria; epsilon subdivision; Helicobacter group;				
OC	Helicobacter				
OX	NCBI_TaxID=210;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=26695 / ATCC 700392;				
RX	MEDLINE=97394467; PubMed=9252185;				
RA	Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,				
RA	Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,				
RA	Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,				
RA	Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,				
RA	McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,				
RA	Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,				
RA	Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,				
RA	Hayes W.S., Borodovsky M., Karpi P.D., Smith H.O., Fraser C.M.,				
RA	Venter J.C.;				
RT	"The complete genome sequence of the gastric pathogen Helicobacter				
RT	pylori.";				
RL	Nature 388:539-547(1997).				
CC	-1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.				
DR	EMBL; AE000610; AAC0063.1; -				
DR	MEROPS; S01.273; -				
DR	TIGR; HP1019; -				
DR	InterPro; IPR001478; PDZ.				
DR	InterPro; IPR001940; Protease2C.				
DR	InterPro; IPR001254; Ser_protease_Try.				
DR	Pfam; PF00595; PDZ; 1.				
DR	Pfam; PF00089; trypsin; 1.				
DR	PRINTS; PR00834; PROTEASES2C.				

```
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
KW Hydrolase: Hypothetical protein; Protease; Serine protease;
KW Complete proteome.
SQ SEQUENCE 443 AA; 47983 MW; 03FDDBD72CF31EE2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 387; DB 16; Length 443;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIPKMERALSGVYISKDGIYVNNHVIDGADTKVTPGNSKEYSATLVGTDSIDL 60
DB 57 MIPKMERALSGVYISKDGIYVNNHVIDGADTKVTPGNSKEYSATLVGTDSIDL 116
QY 61 AVIRITKDNLPITKFSNDSNDISVDGLVFAIGNPFGVGSVTOGIYSALNKGIGINSYEN 120
DB 117 AVIRITKDNLPITKFSNDSNDISVDGLVFAIGNPFGVGSVTOGIYSALNKGIGINSYEN 176
QY 121 FIQTASINPGSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSNMVKDVTVOLIKT 180
DB 177 FIQTASINPGSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSNMVKDVTVOLIKT 236
QY 181 GKIERGYLVGLQDLSGDLQNSYDNKEGAVISVEKDSPAKKAGILVWDLITEVNGKKYK 240
DB 237 GKIERGYLVGLQDLSGDLQNSYDNKEGAVISVEKDSPAKKAGILVWDLITEVNGKKYK 296
QY 241 NTNELNRLIGSMLPNQVTLKVIROKKEKRAFTLTLAERKNPNKKEITISAONGAOGOLNGL 300
DB 297 NTNELNRLIGSMLPNQVTLKVIROKKEKRAFTLTLAERKNPNKKEITISAONGAOGOLNGL 356
QY 301 QVEDLTQETKSRMLSDDDVQGVLSQVNSNENSPAQAGFRQGNIIITKIEEVEVKSADFNNH 360
DB 357 QVEDLTQETKSRMLSDDDVQGVLSQVNSNENSPAQAGFRQGNIIITKIEEVEVKSADFNNH 416
QY 361 ALEKYKPKRFVLVDLNGYRIILVK 387
DB 417 ALEKYKPKRFVLVDLNGYRIILVK 443

RESULT 2
Q92M18 PRELIMINARY; PRT; 476 AA.
ID Q92M18;
AC Q92M18;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Protease DO.
GN HTRA OR JHP0405.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Molr D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.
DR EMBL: AE001474; AAD05980.1; -.
DR MEROPS: S01.273; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.

Query Match
Best Local Similarity 100.0%; Score 387; DB 16; Length 443;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIPKMERALSGVYISKDGIYVNNHVIDGADTKVTPGNSKEYSATLVGTDSIDL 60
DB 57 MIPKMERALSGVYISKDGIYVNNHVIDGADTKVTPGNSKEYSATLVGTDSIDL 116
QY 61 AVIRITKDNLPITKFSNDSNDISVDGLVFAIGNPFGVGSVTOGIYSALNKGIGINSYEN 120
DB 117 AVIRITKDNLPITKFSNDSNDISVDGLVFAIGNPFGVGSVTOGIYSALNKGIGINSYEN 176
QY 121 FIQTASINPGSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSNMVKDVTVOLIKT 180
DB 177 FIQTASINPGSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSNMVKDVTVOLIKT 236
QY 181 GKIERGYLVGLQDLSGDLQNSYDNKEGAVISVEKDSPAKKAGILVWDLITEVNGKKYK 240
DB 237 GKIERGYLVGLQDLSGDLQNSYDNKEGAVISVEKDSPAKKAGILVWDLITEVNGKKYK 296
QY 241 NTNELNRLIGSMLPNQVTLKVIROKKEKRAFTLTLAERKNPNKKEITISAONGAOGOLNGL 300
DB 297 NTNELNRLIGSMLPNQVTLKVIROKKEKRAFTLTLAERKNPNKKEITISAONGAOGOLNGL 356
QY 301 QVEDLTQETKSRMLSDDDVQGVLSQVNSNENSPAQAGFRQGNIIITKIEEVEVKSADFNNH 360
DB 357 QVEDLTQETKSRMLSDDDVQGVLSQVNSNENSPAQAGFRQGNIIITKIEEVEVKSADFNNH 416
QY 361 ALEKYKPKRFVLVDLNGYRIILVK 387
DB 417 ALEKYKPKRFVLVDLNGYRIILVK 443
```

```
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
KW Hydrolase: Serine protease; Complete proteome.
KW Hydrolase: Serine protease; Complete proteome.
SQ SEQUENCE 476 AA; 51713 MW; 4E38AED839BFE18 CRC64;

Query Match
Best Local Similarity 100.0%; Score 96; DB 16; Length 476;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 VTQLIKTGIERGYLVGLQDLSGDLQNSYDNKEGAVISVEKDSPAKKAGILVWDLITE 233
DB 263 VTQLIKTGIERGYLVGLQDLSGDLQNSYDNKEGAVISVEKDSPAKKAGILVWDLITE 322
QY 234 VNGKKVKNTELRNLIGSMLPNQVTLKVIROKKEKRAFTLTLAERKNPNKKEITISAONGAOGOLNGL 300
DB 323 VNGKKVKNTELRNLIGSMLPNQVTLKVIROKKEKRAFTLTLAERKNPNKKEITISAONGAOGOLNGL 358

RESULT 3
Q46094 PRELIMINARY; PRT; 368 AA.
ID Q46094;
AC Q46094;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Heat shock protein/serine protease (Fragment).
GN HTRA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UAS80;
RX MEDLINE=90384493; PubMed=2402249;
RA Taylor D.E., Hiratsuka K.;
RT "Use of non-radioactive DNA probes for detection of Campylobacter
RT jejuni and Campylobacter coli in stool specimens.";
RL Mol. Cell. Probes 4:261-271(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=UAS80;
RX STRAIN=UAS80;
RA Hiratsuka K.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U27271; AAA68943.1; -.
DR MEROPS: S01.273; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
SQ SEQUENCE 368 AA; 39491 MW; CFA14B08EE49EF70 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 368;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 YENFIQTDSINPGSGGAL 137
DB 106 YENFIQTDSINPGSGGAL 125

RESULT 4
Q46120 PRELIMINARY; PRT; 472 AA.
ID Q46120
AC Q46120;
```

01-NOV-1996 (TrEMBLrel. 01, Created)
01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
Serine protease.
HTRA.
Campylobacter jejuni.
Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
Campylobacter.
NCBI_TaxID=197;
[1]
SEQUENCE FROM N.A.
STRAIN-81116;
Henderson J., Wood A.C., Emery M.J., Wren B.W., Kettle J.;
"Characterisation of a Campylobacter jejuni high-temperature-
requirement A (htrA) isogenic mutant."
Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
EMBL: X82628; CAA57948.1;
MEROPS: S01.273;
InterPro: IPR001478; PDZ.
InterPro: IPR001940; Protease2C.
InterPro: IPR001254; Ser_protease_Try.
Pfam: PF00595; PDZ; 2.
Pfam: PF00089; trypsin; 1.
PRINTS: PR00834; PROTEASES2C.
SMART: SM00228; PDZ; 2.
PROSITE: PS50106; PDZ; 1.
PROSITE: PS50240; TRYPSIN_DOM; 1.
Hydrolase; Protease; Serine protease.
KW
SEQUENCE 472 AA; 50941 MW; 0C16E381E97424D1 CRC64;
Query Match 5.2%; Score 20; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 118 YENFIQTDSINPGNSGGAL 137
|
Db 210 YENFIQTDSINPGNSGGAL 229
|
RESULT 5
Q9PN69 PRELIMINARY; PRT; 472 AA.
ID Q9PN69
AC Q9PN69;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Serine protease (protease DO) (EC 3.4.21.).
GN HTRA OR CUL228C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NCTC 11168;
RA MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
Whitehead S., Barrett B.G.;
"The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- STIMULATORY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: ALI39077; CAB73482.1;
DR MEROPS: S01.273;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
01-NOV-1996 (TrEMBLrel. 01, Created)
01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
Serine protease.
HTRA.
Campylobacter jejuni.
Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
Campylobacter.
NCBI_TaxID=197;
[1]
SEQUENCE FROM N.A.
STRAIN-81116;
Henderson J., Wood A.C., Emery M.J., Wren B.W., Kettle J.;
"Characterisation of a Campylobacter jejuni high-temperature-
requirement A (htrA) isogenic mutant."
Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
EMBL: X82628; CAA57948.1;
MEROPS: S01.273;
InterPro: IPR001478; PDZ.
InterPro: IPR001940; Protease2C.
InterPro: IPR001254; Ser_protease_Try.
Pfam: PF00595; PDZ; 2.
Pfam: PF00089; trypsin; 1.
PRINTS: PR00834; PROTEASES2C.
SMART: SM00228; PDZ; 2.
PROSITE: PS50106; PDZ; 1.
PROSITE: PS50240; TRYPSIN_DOM; 1.
Hydrolase; Protease; Serine protease.
KW
SEQUENCE 472 AA; 50941 MW; 0C16E381E97424D1 CRC64;
Query Match 5.2%; Score 20; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 118 YENFIQTDSINPGNSGGAL 137
|
Db 210 YENFIQTDSINPGNSGGAL 229
|
RESULT 5
Q9PN69 PRELIMINARY; PRT; 472 AA.
ID Q9PN69
AC Q9PN69;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Serine protease (protease DO) (EC 3.4.21.).
GN HTRA OR CUL228C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NCTC 11168;
RA MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
Whitehead S., Barrett B.G.;
"The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- STIMULATORY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: ALI39077; CAB73482.1;
DR MEROPS: S01.273;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.

DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Complete proteome.
SQ SEQUENCE 472 AA; 51006 MW; DCE4B51540E9894B CRC64;
Query Match 5.2%; Score 20; DB 16; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 118 YENFIQTDSINPGNSGGAL 137
|
Db 210 YENFIQTDSINPGNSGGAL 229
|
RESULT 6
Q31388 PRELIMINARY; PRT; 371 AA.
ID Q31388
AC Q31388;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE DegP protein.
DE DegP.
GN DegP.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98114461; PubMed=9446679;
RA Narberhaus F., Weiglhofer W., Fischer H.M., Hennecke H.;
"Identification of the Bradyrhizobium japonicum degP gene as part of
an operon containing small heat shock protein genes."
RL Arch. Microbiol. 169:89-97(1998).
DR EMBL: Y13616; CAA73938.1;
DR HSSP; P31016; 1BFE.
DR MEROPS; S01.274;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 371 AA; 39108 MW; 3CE1C77E17B86CDC CRC64;
Query Match 4.7%; Score 18; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 121 FIQTDSINPGNSGGALI 138
|
Db 204 FIQTDSINPGNSGGALI 221
|
RESULT 7
Q9PGL3 PRELIMINARY; PRT; 481 AA.
ID Q9PGL3
AC Q9PGL3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Heat shock protein.
GN XFO285.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]

RP SEQUENCE FROM N.A.
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Baptista M.,
RA Alvares A.R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Colombo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Martins A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins A.E.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.N., Miracca E.C., Miyaki C.Y., Montello-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peikoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Sanceli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,
RT "The genome sequence of the plant pathogen Xylella fastidiosa";
RL Nature 406:151-159(2000).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AE003881; AAF83098.1; -;
DR MEROPS; S01.274; -;
DR InterPro; IPR001478; PDZ;
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001234; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Complete proteome.
SQ SEQUENCE 481 AA; 50536 MW; E90D291491D8EC18 CRC64;

Query Match 4.7%; Score 18; DB 16; Length 481;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 120 NFQTDASINPGNSGGAL 137
|||||
DB 222 NFQTDASINPGNSGGAL 239
|||||

RESULT 8
Q8UE46 PRELIMINARY; PRT; 495 AA.
AC Q8UE46;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Serine protease.
HTRA OR ATU1915 OR AGR_C3507.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,

RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gilllet W., Grant C.,
RA Kutayav N.T., Levy R., Li M.-J., McLelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Doilan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ourullo B., Goldman B.S., Cao Y., Askenzi M., Halling C., Mullin L.,
RA Houmle K., Gordon J., Vaudin M., Doughty D., Iatchouk O., Epp A., Liu F.,
RA Wollman C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58";
RL Science 294:2323-2328(2001).
DR EMBL; AE009145; AAL42911.1; ALT_INIT.
DR EMBL; AE008110; AAK87675.1; -;
KW Protease; Complete proteome.
SQ SEQUENCE 495 AA; 52681 MW; 36B3887F948F9665 CRC64;

Query Match 4.4%; Score 17; DB 16; Length 495;
Best Local Similarity 100.0%; Pred. No. 2.4e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 FIOTDASINPGNSGGAL 137
|||||
DB 230 FIOTDASINPGNSGGAL 246
|||||

RESULT 9
Q97LU1 PRELIMINARY; PRT; 348 AA.
AC Q97LU1;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Serine protease Do (heat-shock protein).
GN CAC0463.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007561; AAK78443.1; -;
DR InterPro; IPR001478; PDZ;
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR000126; Ser_protease_V8.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Protease; Complete proteome.

```
SQ SEQUENCE 348 AA; 36310 MW; 1485D7E9D6FDB2F3 CRC64;
Query Match 3.6%; Score 14; DB 16; Length 348;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 IOTDASINPGNSGG 135
DB 194 IOTDASINPGNSGG 207

RESULT 10
Q9R6V1 PRELIMINARY; PRT; 367 AA.
AC Q8R6V1;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Trypsin-like serine protease, typically periplasmic, contain
DE C-terminal PDZ domain.
GN DEQ3 OR TTE2683.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RA "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013208; AA25802.1; -.
KW Protease; Complete proteome.
SQ SEQUENCE 367 AA; 39991 MW; A99CB36C029E1AF2 CRC64;

Query Match 3.6%; Score 14; DB 16; Length 367;
Best Local Similarity 100.0%; Pred. No. 3.5e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 IOTDASINPGNSGG 135
DB 212 IOTDASINPGNSGG 225

RESULT 11
Q93J30 PRELIMINARY; PRT; 519 AA.
AC Q93J30;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative protease.
GN SC03977 OR SCBAC25E3.14.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Collins M., Harris D.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
```

```
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL596251; CAC44701.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001234; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR KW Hydrolase; Protease; trypsin; 1.
SQ SEQUENCE 519 AA; 50327 MW; 5CB8D5F0CC19E428 CRC64;

Query Match 3.4%; Score 13; DB 16; Length 519;
Best Local Similarity 100.0%; Pred. No. 6.1e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 QTDASINPGNSGG 135
DB 352 QTDASINPGNSGG 364

RESULT 12
Q99TD6 PRELIMINARY; PRT; 424 AA.
AC Q99TD6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SAV1728.
GN SAV1728 OR SA1545.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Kanamori M., Matsumaru H., Murayama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirukawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogisawa N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AP003363; BAB57890.1; -.
DR EMBL; AP003134; BAB42817.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001178; PDZ.
DR InterPro; IPR001340; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try..
```

DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Hypothetical protein; Complete proteome.
SQ SEQUENCE 424 AA; 45803 MW; 888ADBA9A6E97948 CRC64;

Query Match 3.1%; Score 12; DB 16; Length 424;
Best Local Similarity 100.0%; Pred. No. 0.00062;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 YIVTNNHVIDGA 33
IIIIIIIIIIIIIIIIIIII
Db 138 YIVTNNHVIDGA 149

RESULT 13
O68197 PRELIMINARY; PRT; 437 AA.
AC O68197;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Htra (Fragment).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NTHI 12;
RA Loosmore S.M., Yang Y.P., Oomen R., Shortreed J.M., Coleman D.C., Klein M.H.;
RT "The Haemophilus influenzae Htra protein is a protective antigen.";
RL Infect. Immun. 66:899-906(1998).
DR EMBL: AF018151; AAC38202.1; -.
DR MEROPS: S01.274; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 437 AA; 46376 MW; 27AE2E9F27BE5F6C CRC64;

Query Match 3.1%; Score 12; DB 2; Length 437;
Best Local Similarity 100.0%; Pred. No. 0.00064;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 TNNHVIDGADKI 36
IIIIIIIIIIIIIIIIIIII
Db 88 TNNHVIDGADKI 99

RESULT 14
Q56885 PRELIMINARY; PRT; 464 AA.
AC Q56885;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Htra protein (Fragment).
GN HTRA.
OS Yersinia enterocolitica.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-8081V;
RX MEDLINE-962339021; PubMed-8675311;
RA Li S.R., Dorrell N., Everest P.H., Dougan G., Wren B.W.;
RT "Construction and characterisation of a Yersinia enterocolitica 0:8 high temperature requirement (htra) isogenic mutant.";
RL Infect. Immun. 64:2088-2094(1996).
DR EMBL: X94153; CAA63869.1; -.
DR HSP: P29476; IQAV.
DR MEROPS: S01.274; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000126; Ser_proteas_V8.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 464 AA; 48209 MW; E5D96BF90CFD6542 CRC64;

Query Match 3.1%; Score 12; DB 2; Length 464;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNVVK 171
IIIIIIIIIIIIIIIIIIII
Db 248 GIGFAIPSNVVK 259

RESULT 15
Q44596 PRELIMINARY; PRT; 474 AA.
AC Q44596;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Htra-like protein.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=235;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2308;
RX MEDLINE-95165990; PubMed-7861951;
RA Tatum F.M., Cheville N.F., Morfitt D.;
RT "Cloning, characterization and construction of htra and htra-like mutants of Brucella abortus and their survival in BALB/c mice.";
RL Microb. Pathog. 17:23-36(1994).
DR EMBL: U07351; AAA70163.1; -.
DR MEROPS: S01.273; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 474 AA; 50159 MW; F74AC9960C4EF9FB CRC64;

Query Match 3.1%; Score 12; DB 2; Length 474;

Best Local Similarity 100.0%; Pred. No. 0.00069;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 INPNSGGALID 139

|||||

Db 215 INPNSGGALID 226

Search completed: November 18, 2002, 11:35:08
Job time : 37 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 18, 2002, 11:27:10 ; Search time 35 Seconds
(without alignments)
2278.292 Million cell updates/sec

Title: US-09-895-913A-120
Perfect score: 1938
Sequence: 1 MIPKERMALSGVILSKD.....KPKRFLVLDLNGGYRIILVK 387

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1938	100.0	443	16	O25663 helicobacte
2	1914	98.8	476	16	Q9ZM18 helicobacte
3	937	48.3	472	16	Q9PN69 campylobacte
4	934.5	48.2	472	2	Q46120 campylobact
5	921	47.5	368	2	Q46094 campylobact
6	767	39.6	456	16	Q9KUF5 vibrio chol
7	753.5	38.9	453	2	Q9FD11 aeromonas h
8	717.5	37.0	475	16	Q8Z9B0 salmonella
9	710	36.6	453	16	O67436 aquifex aeo
10	707.5	36.5	474	16	Q8YHL4 brucella me
11	706.5	36.5	474	2	Q9ALS1 pseudomonas
12	706.5	36.5	474	16	O57155 pseudomonas
13	706.5	36.5	477	2	Q8RSS1 klebsiella
14	704.5	36.4	491	2	Q9LBR0 shigella so
15	703.5	36.3	474	2	Q44596 brucella ab
16	698	36.0	457	16	Q8ZB58 yersinia pe

17	693.5	35.8	513	16	Q98CS8 rhizobium l
18	692.5	35.7	523	16	Q8UGQ8 agrobacteri
19	691.5	35.7	403	16	Q8XV99 ralstonia s
20	679.5	35.1	513	16	Q8YQ32 brucella me
21	675.5	34.9	481	2	Q9AQD1 pseudomonas
22	671	34.6	437	2	O68197 haemophilus
23	670.5	34.6	478	2	P74978 yersinia en
24	667.5	34.4	481	16	Q8ZBM6 yersinia pe
25	666	34.4	463	2	O68198 haemophilus
26	664	34.3	459	16	Q8CMS7 pasteurella
27	655	33.8	455	16	Q8Z3E6 salmonella
28	653	33.7	514	16	Q9PBA3 xyella fas
29	652	33.6	455	16	Q8X9F1 escherichia
30	651	33.6	455	16	Q8ZLQ1 salmonella
31	651	33.6	514	16	Q8UDS7 agrobacteri
32	646	33.3	500	2	Q44652 brucella ab
33	645	33.3	464	2	Q56885 yersinia en
34	639	33.0	495	16	Q8UE46 agrobacteri
35	635.5	32.8	481	16	Q9PGL3 xyella fas
36	634	32.7	516	16	Q985F9 rhizobium l
37	627	32.4	511	2	Q8RTK2 xanthomonas
38	625	32.2	490	16	Q8XPT5 ralstonia s
39	621.5	32.1	389	16	Q9HYX1 pseudomonas
40	619.5	32.0	505	16	Q8Y016 ralstonia s
41	618.5	31.9	371	2	O31388 bradyrhizob
42	611.5	31.6	504	16	Q98KJ1 rhizobium l
43	610.5	31.5	428	16	Q98N31 rhizobium l
44	604.5	31.2	500	2	Q9KJN6 myxococcus
45	602.5	31.1	503	16	Q9Z6C8 rhizobium m

ALIGNMENTS

RESULT 1

O25663	ID	O25663	PRELIMINARY;	PRT;	443 AA.
AC	O25663;				
DT	01-JAN-1998	(TRENBLrel. 05, Created)			
DT	01-JAN-1998	(TRENBLrel. 05, Last sequence update)			
DT	01-MAR-2002	(TRENBLrel. 20, Last annotation update)			
DE	Serine protease (HTRA).				
GN	HP1019.				
OS	Helicobacter pylori (Campylobacter pylori).				
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;				
OC	Helicobacter.				
OX	NCBI_TaxID=210;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=26695 / A/CC 700392;				
RX	MEDLINE=97394467; PubMed=9252185;				
RA	Tomb J.-F., Whit; O., Kerlavage A.R., Clayton R.A., Sutton G.G.,				
RA	Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,				
RA	Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,				
RA	Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,				
RA	McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,				
RA	Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,				
RA	Hayes W.S., Borodovsky M., Fujii C., Bowman C., Watthey L., Wallin E.,				
RA	Venter J.C.;				
RT	"The complete genome sequence of the gastric pathogen Helicobacter				
RL	pylori.";				
CC	Nature 388:539-547(1997).				
DR	-1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.				
DR	EMBL; AE000610; AAD08063.1; -				
DR	MEROPS; S01.273; -				
DR	TIGR; HP1019; -				
DR	InterPro; IPR001478; PDZ.				
DR	InterPro; IPR001940; Protease2C.				
DR	InterPro; IPR001254; Ser_protease_Try.				
DR	Pfam; PF00595; PDZ; 1.				
DR	Pfam; PF00089; trypsin; 1.				
DR	PRINTS; PR00834; PROTEASES2C.				

DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
KW Hydrolase; Hypothetical protein; Protease; Serine protease;
KW Complete proteome.
SQ SEQUENCE 443 AA; 47983 MW; 03FDDBD72CF31EE2 CRC64;

Query Match 100.0%; Score 1938; DB 16; Length 443;
Best Local Similarity 100.0%; Pred. No. 1.3e-98;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIPKERMERALSGVVISKDGIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSESDL 60
DB 57 MIPKERMERALSGVVISKDGIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSESDL 116
QY 61 AVIRITKDNLPITKFSDSNDISVGDVFAIGNPFGVSVTQGIIVSALNKGIGINSYEN 120
DB 117 AVIRITKDNLPITKFSDSNDISVGDVFAIGNPFGVSVTQGIIVSALNKGIGINSYEN 176
QY 121 FIOTDASINPGSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSNMVKDTVTQLIKT 180
DB 177 FIOTDASINPGSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSNMVKDTVTQLIKT 236
QY 181 GKIERGYLVGLQDLSDGLQNSYDNKEGAVVISVEKDSPPAKKAGILVWDLITEVNGKKVK 240
DB 237 GKIERGYLVGLQDLSDGLQNSYDNKEGAVVISVEKDSPPAKKAGILVWDLITEVNGKKVK 296
QY 241 NTNELNRLNGSLMPLNQRVTLKVRDKKERAFITLTLAERKNPNKKTETISAQNGAQQQLNGL 300
DB 297 NTNELNRLNGSLMPLNQRVTLKVRDKKERAFITLTLAERKNPNKKTETISAQNGAQQQLNGL 356
QY 301 QVEDLTQETKRSMRLSDVGVVLVSQVNSNPAPQAGFRQGNITITKIEEVEKSVADFNH 360
DB 357 QVEDLTQETKRSMRLSDVGVVLVSQVNSNPAPQAGFRQGNITITKIEEVEKSVADFNH 416
QY 361 ALEKYGKPKRFLVLDLNOGYRIILVK 387
DB 417 ALEKYGKPKRFLVLDLNOGYRIILVK 443

RESULT 2
Q92M18
ID Q92M18 PRELIMINARY; PRT; 476 AA.
AC Q92M18;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Protease DO.
GN HTRA OR JHP0405.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dalg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tumino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.
DR EMBL: AF001474; RAD05980.1; -;
DR MEROPS: S01.273; -;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.

DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
KW Hydrolase; Serine protease; Complete proteome.
SQ SEQUENCE 476 AA; 51713 MW; 4E338AED839BF18 CRC64;

Query Match 98.8%; Score 1914; DB 16; Length 476;
Best Local Similarity 98.7%; Pred. No. 2.9e-97;
Matches 382; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MIPKERMERALSGVVISKDGIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSESDL 60
DB 90 MIPKERMERALSGVVISKDGIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSESDL 149
QY 61 AVIRITKDNLPITKFSDSNDISVGDVFAIGNPFGVSVTQGIIVSALNKGIGINSYEN 120
DB 150 AVIRITKDNLPITKFSDSNDISVGDVFAIGNPFGVSVTQGIIVSALNKGIGINSYEN 209
QY 121 FIOTDASINPGSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSNMVKDTVTQLIKT 180
DB 210 FIOTDASINPGSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSNMVKDTVTQLIKT 269
QY 181 GKIERGYLVGLQDLSDGLQNSYDNKEGAVVISVEKDSPPAKKAGILVWDLITEVNGKKVK 240
DB 270 GKIERGYLVGLQDLSDGLQNSYDNKEGAVVISVEKDSPPAKKAGILVWDLITEVNGKKVK 329
QY 241 NTNELNRLNGSLMPLNQRVTLKVRDKKERAFITLTLAERKNPNKKTETISAQNGAQQQLNGL 300
DB 330 NTNELNRLNGSLMPLNQRVTLKVRDKKERAFITLTLAERKNPNKKTETISAQNGAQQQLNGL 389
QY 301 QVEDLTQETKRSMRLSDVGVVLVSQVNSNPAPQAGFRQGNITITKIEEVEKSVADFNH 360
DB 390 QVEDLTQETKRSMRLSDVGVVLVSQVNSNPAPQAGFRQGNITITKIEEVEKSVADFNH 449
QY 361 ALEKYGKPKRFLVLDLNOGYRIILVK 387
DB 450 ALEKYGKPKRFLVLDLNOGYRIILVK 476

RESULT 3
Q9PN69
ID Q9PN69 PRELIMINARY; PRT; 472 AA.
AC Q9PN69;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Serine protease (protease DO) (EC 3.4.21.).
GN HTRA OR CJ1228C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.
DR EMBL: AL139077; CAB73482.1; -;
DR MEROPS: S01.273; -;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.

DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Complete proteome.
SQ SEQUENCE 472 AA; 51006 MW; DCE4B51540E9894B CRC64;

Query Match 48.3%; Score 937; DB 16; Length 472;
Best Local Similarity 51.3%; Pred. No. 9.4e-44;
Matches 195; Conservative 69; Mismatches 102; Indels 14; Gaps 5;

QY 10 ALGSGVIISKDGYIVTNHHVIDGADKIKVTIPGSNKEYSATLVGTDSDSLAVIRITKDN 69
DB 102 SLGSGVIISKDGYIVTNHHVIDDADTIIVNLPGSDEYKAKLIGDKPTDLAVIKIENN 161
QY 70 LPTIKFSNDISVGDVFAIGNPFGVGSVTQGIIVSALNKGSGIGINSYENFIQTDA 129
DB 162 LSATFTNSDDLLMEGDVVFALGNPFGVGSVTSGIISALNKNIGLNQYENFIQTDA 221
QY 130 PNSSGALIDSRGVLGINTAIISKTGNGHIGFAIPSNMVKDVTQLIKTKIERGYLG 189
DB 222 PNSSGALVDSRGYLVGINSAIISLRGGNGGIGFAIPSNMVKDIAKLTIEKIDRGFLG 281
QY 190 VGLQDLSGDLQNSYDNKEGAVVISVEKDSPAKKAGILVWDLITEVNGKVKNTNELRLI 249
DB 282 VTILALQDGTKKAYKNQEGALITDVQKSSADEAGLKRGLDVTYKNDVKVIRPIDLKNYI 341
QY 250 GMLPNQRVTLKVID--KERAFTLTIAERKNPNKKTISQAQGAQGL-NGLOVEDLT 306
DB 342 GTLEIGQKISLSYERDGENKQASFILK-GEKENP-----KGVQSDLDIGLSRLNLD 391
QY 307 QETKSMRLSDVQGVLSQVNSPAGQAFRQGNITKIEEVEKSVADPNHALEYK 366
DB 392 PRLKDLQTPKDVNGVLVDSVKEKSGKNSGFQEGDIIIGVQSQSEIKNLKDLQAL-KQV 450
QY 367 GKPKFLVLDLNGVRIILV 386
DB 451 NKKEFTKVVYRNGPATLLV 470

RESULT 4
Q46120
ID Q46120 PRELIMINARY; PRT; 472 AA.
AC Q46120;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Serine protease.
GN HTRA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81116;
RA Henderson J., Wood A.C., Emery M.J., Wren B.W., Ketley J.;
RT "Characterisation of a Campylobacter jejuni high-temperature-
requirement A (htrA) isogenic mutant."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82628; CAA57948.1; -;
DR MEROPS; S01.273; -;
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 472 AA; 50941 MW; 0C16E381E97424D1 CRC64;

Query Match 48.2%; Score 934.5; DB 2; Length 472;
Best Local Similarity 52.2%; Pred. No. 1.3e-43;
Matches 189; Conservative 68; Mismatches 92; Indels 13; Gaps 4;

QY 10 ALGSGVIISKDGYIVTNHHVIDGADKIKVTIPGSNKEYSATLVGTDSDSLAVIRITKDN 69
DB 102 SLGSGVIISKDGYIVTNHHVIDDADTIIVNLPGSDEYKAKLIGDKPTDLAVIKIENN 161
QY 70 LPTIKFSNDISVGDVFAIGNPFGVGSVTQGIIVSALNKGSGIGINSYENFIQTDA 129
DB 162 LSATFTNSDDLLMEGDVVFALGNPFGVGSVTSGIISALNKNIGLNQYENFIQTDA 221
QY 130 PNSSGALIDSRGVLGINTAIISKTGNGHIGFAIPSNMVKDVTQLIKTKIERGYLG 189
DB 222 PNSSGALVDSRGYLVGINSAIISLRGGNGGIGFAIPSNMVKDIAKLTIEKIDRGFLG 281
QY 190 VGLQDLSGDLQNSYDNKEGAVVISVEKDSPAKKAGILVWDLITEVNGKVKNTNELRLI 249
DB 282 ASILALQDGTKKAYKNQEGALITDVQKSSADEAGLKRGLDVTYKNDVKVIRPIDLKNYI 341
QY 250 GMLPNQRVTLKVID--KERAFTLTIAERKNPNKKTISQAQGAQGL-NGLOVEDLT 306
DB 342 GTLEIGQKISLSYERDGENKQASFILK-GEKENP-----KGVQSDLDIGLSRLNLD 391
QY 307 QETKSMRLSDVQGVLSQVNSPAGQAFRQGNITKIEEVEKSVADPNHALEYK 366
DB 392 PRLKDLQTPKDVNGVLVDSVKEKSGKNSGFQEGDIIIGVQSQSEIKNLKDLQAL-KQV 451
QY 367 GK 368
DB 452 KK 453

RESULT 5
Q46094
ID Q46094 PRELIMINARY; PRT; 368 AA.
AC Q46094;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Heat shock protein/serine protease (Fragment).
GN HTRA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA580;
RX MEDLINE-90384493; PubMed-2402249;
RA Taylor D.E., Hiratsuka K.;
RT "Use of non-radiocative DNA probes for detection of Campylobacter
jejuni and Campylobacter coli in stool specimens."
RL Mol. Cell. Probes 4:261-271(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UA580;
RA Hiratsuka K.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U27271; AAA68943.1; -;
DR MEROPS; S01.273; -;
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001354; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
SQ SEQUENCE 368 AA; 39491 MW; CFA14B08EE49EF70 CRC64;

```
Query Match 47.5%; Score 921; DB 2; Length 360;
Best Local Similarity 50.9%; Pred. No. 5.1e-43;
Matches 192; Conservative 69; Mismatches 102; Indels 14; Gaps 5;

QY 13 SGVLIISKDGIVTNNHVIDGADKIKVTPGSKKEYSATLVGTDSGLAVIRITKONLPT 72
DB 1 SGVLIISKDGIVTNNHVIDGADKIKVTPGSKKEYSATLVGTDSGLAVIRITKONLPT 72
QY 73 IKFSDSNISVGLDFAIGNPFGVSVTQGIYSALNKGISGINSYENFIQTDSINPGN 132
DB 61 ITFNSDDLMEGDVFAIGNPFGVSVTQGIYSALNKGISGINSYENFIQTDSINPGN 120
QY 133 SGGALLDSRGGLVGINTAIISKGTGGNHGIGFAIPSNMVKDVTVTLKTKIERGYLGVL 192
DB 121 SGGALVDSRGYLVGINSAILSRGGNGGIGFAIPSNMVKDIAKKIEKGKIDRGFLGVTI 180
QY 193 QDLSGDLNSYDNKEGAVNISVEKDSPPAKKAGILVWDLITEYNGKKVKNTELRNLIGSM 252
DB 181 LALQGNTKKAYKNOEGALITDVQGSADPAQLKRGDLVTVKNDKVIKSPIDLKNTIGTL 240
QY 253 LPNQRVTLKVIRO--KKERAFTTLAERKNPNKKTETISAQNGAOGOL--NGLOVEDLTQET 309
DB 241 EIQKISLSVERDGENKQASFILK--GEKENP-----KGVSOLDLIGLSLRNLDPL 290
QY 310 KSMRLSDNVQGVVLVSQVNSNPAEQAGFRQGNITTKIEVEVKSVDNPHALEKYGKXP 369
DB 291 KDRLOIPKDVNGVLVSHVREKSKGKSGFQEGDIIGVGSGEIKNEKLDLEQAL--KQVKK 349
QY 370 KRLVLDLNGQYRIILV 386
DB 350 EFTKVCYVRNGFATLLV 366

RESULT 6
Q9KUF5 PRELIMINARY; PRT; 456 AA.
AC Q9KUF5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE Protease DO.
GN VC0566.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AF004142; AAF93734.1; -.
DR MEROPS; S01.274; -.
DR TIGR; VC0566; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00089; trypsin; 1.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.

KW Hydrolase; Serine protease; Complete proteome.
SQ SEQUENCE 456 AA; 48366 MW; B4D3912BE87B0AEE CRC64;

Query Match 39.6%; Score 767; DB 16; Length 456;
Best Local Similarity 43.6%; Pred. No. 1.8e-34;
Matches 168; Conservative 83; Mismatches 112; Indels 22; Gaps 8;

QY 4 KERMERALSGVLIISKD-GYIVTNNHVIDGADKIKVTPGSKKEYSATLVGTDSGLAV 62
DB 85 QERPERGLSGVLIINADKGYVVTNNHYHNGAERIKVKL-YDGRFDEALVGGDEMSDAL 143
QY 63 IRTTK-DNLPITIKFSDSNISVGLDFAIGNPFGVSVTQGIYSALNKGISGINSYENFI 121
DB 144 LKLNKAKNLTEIRIADSDKLVRGDFAVAGNPFGLGTQVTSIGVLSALRSGSLNIENF 203
QY 122 IQTDASINPNSGGLIDSRGGLVGINTAIISKGTGGNHGIGFAIPSNMVKDVTVTLKTK 181
DB 204 IQTDAINSGSGALVNLNGELIGINTAILGPNNGVNGIGFAIPSNMKNLTDQILLEG 263
QY 182 KIERGYLGVLQDLSDGLQNS--YDNKEGAVNISVEKDSPPAKKAGILVWDLITEYNGKKV 239
DB 264 EVKRGMLGVQGGIEITSELADALGYESSKGAFVSVQVVPDAAADKAGIKAGDIITSLNGKKI 323
QY 240 KNTNELNLIGSMPLNQRVTLKVIROKKEKRAFTTLAERKNPNKKTETISAQNGAOGOLNG 299
DB 324 DTFSELRAKVATLGAAGKTTITGLVRDGNQINIDVTILGEQQNAKTK----AESLHQG-LSG 378
QY 300 LOVEDLTQETKSRMRLSDNVQGVVLVSQVNSNPAEQAGFRQGNITTKIEVEVKSVDN 359
DB 379 AELSNTD-----SDPIQGVKVTVEVQKSGSAESYQLOKDDIIIGVNRKRKNIAELR 430
QY 360 HALEYKGRPKRFLVLDLNGQYRII 384
DB 431 AIME----KSPNIALNIQGRERTL 451

RESULT 7
Q9FD11 PRELIMINARY; PRT; 453 AA.
AC Q9FD11;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE HtrA-like serine protease.
GN PRTS1.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OX NCBI_TaxID=644;
RN [1]
RC STRAIN=CKH-29;
RX SEQUENCE FROM N.A.
RA Lin T.-N., Lin T.-J., Liou C.-M.;
RT "Aeromonas hydrophila strain CKH-29 prts1 gene complete cds."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF293977; AAG03073.1; -.
DR MEROPS; S01.274; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00089; trypsin; 1.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 453 AA; 47828 MW; BD55CE2B844E5148 CRC64;

Query Match 38.9%; Score 753.5; DB 2; Length 453;
Best Local Similarity 43.0%; Pred. No. 1e-33;
Matches 166; Conservative 85; Mismatches 108; Indels 27; Gaps 9;
```



```
DB 268 SRGVLGVVQIEVKNKDLAESFGLDKPAGALVAQLVEDGPAAGGLQGVGVVLSLNGQSINE 327
QY 242 TNELNLNLSGMLPNORVTLKVRDKKERAFTLTTLAERKNPNKKEITSAQN--GAQGQLN- 298
DB 328 SADLPHLVGNMKPGDKINLVDVIRNGQRK--SUSMVGSLPDDDEIASMGAPGAERSNR 385
QY 299 -GLOVEDLTQETKRSMLSDDDVQGVLSQVNSNSPAEQAFRQGNITIKIEVEVKS-- 355
DB 386 LGVTVAADLTAEQKSL----DIQGGVWIKVQDGAAGVIGLPGDVIHLDNKAFTSTKI 441
QY 356 -ADFNHALEKYKGPKEFREL 373
DB 442 FADVAKALPKNRSVSMRVL 460

RESULT 12
Q57155 PRELIMINARY; PRT; 474 AA.
AC Q57155;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE MUCD (Serine protease MUCD).
GN MUCD OR PA0766.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=93391358; PubMed=8378309;
RA Martin D.W., Schurr M.J., Mudd M.H., Govan J.R., Holloway B.W.,
RA Deretic V.;
RT "Mechanism of conversion to mucoidy in Pseudomonas aeruginosa
RT infecting cystic fibrosis patients.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8377-8381(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=95286510; PubMed=7768826;
RA Yu H., Schurr M.J., Deretic V.;
RT "Functional equivalence of Escherichia coli sigma E and Pseudomonas
RT aeruginosa AlgU: E. coli rpoE restores mucoidy and reduces sensitivity
RT to reactive oxygen intermediates in algU mutants of P. aeruginosa.";
RL J. Bacteriol. 177:3259-3268(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=96134987; PubMed=8550474;
RA Boucher J.C., Martinez-Salazar J., Schurr M.J., Mudd M.H., Yu H.,
RA Deretic V.;
RT "Two distinct loci affecting conversion to mucoidy in Pseudomonas
RT aeruginosa in cystic fibrosis encode homologs of the serine protease
RT HtrA.";
RL J. Bacteriol. 178:511-523(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; U49151; AAC43718.1; -.
DR EMBL; U32853; AAC43676.1; -.

us-09-895-913a-120_2.rspt
```

```
DR EMBL; AE004511; AA04155.1; -.
DR InterPro; IPR001473; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 474 AA; 50321 MW; 8AB5D2A89867BEBE CRC64;

Query Match 36.5%; Score 706.5; DB 16; Length 474;
Best Local Similarity 41.7%; Pred. No. 4e-31;
Matches 158; Conservative 78; Mismatches 126; Indels 17; Gaps 8;

QY 4 KERMALGSGVVISKDGIVITVNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSGLAVI 63
DB 90 QOREAQSILSGFIIISNDGYILTNHVVADAEILVRL-SDRSEHKAKLTGADPRSDVAVL 148
QY 64 RTKDNLTPIKFSDSNDISVGLVFAIGNPFGVSGVTQGIYSALNKGIGINSYENFTQ 123
DB 149 KIEAKNLPTLKLGDNSKLNKLVGEWVLAIGSPFGFDHVSHTAGIVSAKGRS-LPNESVVPFIQ 207
QY 124 TDASINPGNSGALIDSRGLVGINTAIIISKGGNHGIGFAIPSNMVKDVTVTLQIKTKI 183
DB 208 TDVAINPGNSGGVLLNLOGEVVGINSQIFTRSGGFMGLSFAIPIDVALNVADLKKACKV 267
QY 184 ERGYLGVLQDLISGDLQNSY--DNKEGAVIVSEKDSAPAKKAGILVMDLITEVNGKKVKN 241
DB 268 SRGVLGVVQIEVKNKDLAESFGLDKPAGALVAQLVEDGPAAGGLQGVGVVLSLNGQSINE 327
QY 242 TNELNLNLSGMLPNORVTLKVRDKKERAFTLTTLAERKNPNKKEITSAQN--GAQGQLN- 298
DB 328 SADLPHLVGNMKPGDKINLVDVIRNGQRK--SUSMVGSLPDDDEIASMGAPGAERSNR 385
QY 299 -GLOVEDLTQETKRSMLSDDDVQGVLSQVNSNSPAEQAFRQGNITIKIEVEVKS-- 355
DB 386 LGVTVAADLTAEQKSL----DIQGGVWIKVQDGAAGVIGLPGDVIHLDNKAFTSTKI 441
QY 356 -ADFNHALEKYKGPKEFREL 373
DB 442 FADVAKALPKNRSVSMRVL 460

RESULT 13
Q8RSS1 PRELIMINARY; PRT; 477 AA.
AC Q8RSS1;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE HtrA protein.
GN HTRA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RA Cortes G.;
RT "Role of the htrA gene in Klebsiella pneumoniae virulence.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430233; CAD22887.1; -.
SQ SEQUENCE 477 AA; 49559 MW; 28D222FAE59FCF487 CRC64;

Query Match 36.5%; Score 706.5; DB 2; Length 477;
Best Local Similarity 42.7%; Pred. No. 4e-31;
Matches 160; Conservative 74; Mismatches 116; Indels 25; Gaps 9;

QY 10 ALGSGVII-SKDGIVITVNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSGLAVI-TK 67
||||||| : ||:|||||:| ||| : : ||| | ||:| |
```


QY 240 KNTNELNLIGSMLPNQVTLKVIIRDKKERAFTLTIAERKNPNKKETISAQNGAQOQLNG 299
| : | : : : : | : | : | :
Db 330 DNQDVLGYRLSTAGICKTISVEVMRGNLSLPVKLT--KAPKVQAEFKVIEGDNPEFG 387
QY 300 LOVEDLTQETKRSMLSDDVQGVLYSQVNSPBAEQAGFRQGNIIITKIEEVEVKSVDEN 359
| : | : : : : | : | : | : | : : : | :
Db 388 AAVGDLTASTAAKLRLKRGQGVAVFDVYSGCPAARLGLRSGDIIRSGINGQIRTVDDMT 447
QY 360 HALEKYKG 367
| : | :
Db 448 AVLEAGRG 455

Search completed: November 18, 2002, 11:29:40
Job time : 38 secs


```

SQ SEQUENCE 348 AA; 36310 MW; 1485D7E9D6FDB2F3 CRC64;
Query Match 3.6%; Score 14; DB 16; Length 348;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 IQTASINPGNSGG 135
DB 194 IQTASINPGNSGG 207
|||||
|||||

RESULT 10
Q8R6V1 PRELIMINARY; PRT; 367 AA.
AC Q8R6V1
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Trypsin-like serine protease, typically periplasmic, contain
DE C-terminal PDZ domain.
GN DEG03 OR TTE2683.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013208; AAM25802.1; -.
KW Protease; Complete proteome.
SQ SEQUENCE 367 AA; 39991 MW; A99CB36C029E1AF2 CRC64;

Query Match 3.6%; Score 14; DB 16; Length 367;
Best Local Similarity 100.0%; Pred. No. 3.5e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 IQTASINPGNSGG 135
DB 212 IQTASINPGNSGG 225
|||||
|||||

RESULT 11
Q93J30 PRELIMINARY; PRT; 519 AA.
AC Q93J30
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative protease.
GN SC03977 OR SCBAC25E3.14.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Collins M., Harris D.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);

```

```

RX MEDLINE=97000351; PubMed=8843436;
RA Kedenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL596251; CAC44701.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 519 AA; 50327 MW; 5CB8D5F0CC19E428 CRC64;

Query Match 3.4%; Score 13; DB 16; Length 519;
Best Local Similarity 100.0%; Pred. No. 6.1e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 QTDASINPGNSGG 135
DB 352 QTDASINPGNSGG 364
|||||
|||||

RESULT 12
Q99TD6 PRELIMINARY; PRT; 424 AA.
AC Q99TD6;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical protein SAV1728.
GN SAV1728 OR SA1549.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
CC -1- SIMILARITY: TG SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AP003363; BAB57890.1; -.
DR EMBL; AP003134; BAB42817.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.

```


Query Match 3.18; Score 12; DB 2; Length 474;

H. pfg599 Tomb et al 1997

```
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Hypothetical protein; Protease; Serine protease;
KW Complete proteome.
SQ SEQUENCE 443 AA; 47983 MW; 03FDDBD72CF31EE2 CRC64;

Query Match          100.0%; Score 387; DB 16; Length 443;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIPKMERALGSGVVISKDGIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTDSESDL 60
Db 57 MIPKMERALGSGVVISKDGIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTDSESDL 116

QY 61 AVIRITKDNLPITKFSNDSNDISVGDVFAIGNPFGVGSVTQGIIVSALNKGSGIGINSYEN 120
Db 117 AVIRITKDNLPITKFSNDSNDISVGDVFAIGNPFGVGSVTQGIIVSALNKGSGIGINSYEN 176

QY 121 FIOTDASINPGNSGGALIDSRGLVGINTAIISKTGNGHIGIFAIPSNMVKDTVTOLIKT 180
Db 177 FIOTDASINPGNSGGALIDSRGLVGINTAIISKTGNGHIGIFAIPSNMVKDTVTOLIKT 236

QY 181 GKTERGLVGQLDLSGDLQNSYDNKEGAVVISVEKDSAPKAGILVLDLITEVNGKKYK 240
Db 237 GKTERGLVGQLDLSGDLQNSYDNKEGAVVISVEKDSAPKAGILVLDLITEVNGKKYK 296

QY 241 NTNELRNLTGSMPLPNQVTLKVRDKKERAFITLTAEARNPNKKTISAQNGAQQOLNGL 300
Db 297 NTNELRNLTGSMPLPNQVTLKVRDKKERAFITLTAEARNPNKKTISAQNGAQQOLNGL 356

QY 301 QVEDLTQETKRSNRLSDDDVQGVLSQVNSPNAEQAFRQGNIIITKIEEVEVKSVDNFH 360
Db 357 QVEDLTQETKRSNRLSDDDVQGVLSQVNSPNAEQAFRQGNIIITKIEEVEVKSVDNFH 416

QY 361 ALEKYGKPKRFVLVDLNGYRIILVK 387
Db 417 ALEKYGKPKRFVLVDLNGYRIILVK 443

RESULT 2
Q9ZM18 ID Q9ZM18 PRELIMINARY; PRT; 476 AA.
AC Q9ZM18;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Protease DO.
GN HTRA OR JHP0405.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AE001474; AA005980.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.

Query Match          5.2%; Score 20; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 9.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 YENFIQTDSINFGNSGGAL 137
Db 106 YENFIQTDSINFGNSGGAL 125

RESULT 4
Q46120 ID Q46120 PRELIMINARY; PRT; 472 AA.
AC Q46120;
```

```
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Complete proteome.
SQ SEQUENCE 476 AA; 51713 MW; 4E338AED839BFE18 CRC64;

Query Match          24.8%; Score 96; DB 16; Length 476;
Best Local Similarity 100.0%; Pred. No. 7.5e-96;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 VTQLIKTKIERGYLGVLQDLQSGDLQNSYDNKEGAVVISVEKDSAPKAGILVWDLITE 233
Db 263 VTQLIKTKIERGYLGVLQDLQSGDLQNSYDNKEGAVVISVEKDSAPKAGILVWDLITE 322

QY 234 VNGKKVKNTELNRNLGSLMPLPNQVTLKVIROKKER 269
Db 323 VNGKKVKNTELNRNLGSLMPLPNQVTLKVIROKKER 358
```

Taylor et al

```
RESULT 3
Q46094 ID Q46094 PRELIMINARY; PRT; 368 AA.
AC Q46094;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Heat shock protein/serine protease (Fragment).
GN HTRA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UA580;
RX MEDLINE=90384493; PubMed=2402249;
RA Taylor D.E., Hiratsuka K.;
RT "Use of non-radioactive DNA probes for detection of Campylobacter
RT jejuni and Campylobacter coli in stool specimens.";
RL Mol. Cell. Probes 4:261-271(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=UA580;
RA Hiratsuka K.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U27271; AAA68943.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Protease; Serine protease.
FT NON-TER 1
SQ SEQUENCE 368 AA; 39491 MW; CFA14B08EE49EF70 CRC64;

Query Match          5.2%; Score 20; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 9.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 YENFIQTDSINFGNSGGAL 137
Db 106 YENFIQTDSINFGNSGGAL 125

RESULT 4
Q46120 ID Q46120 PRELIMINARY; PRT; 472 AA.
AC Q46120;
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:29:46 ; Search time 11 Seconds
(without alignments)
529.869 Million cell updates/sec

Title: US-09-895-913A-120
Perfect score: 1938
Sequence: 1 MIPKMERALGSGVLIISKD.....KPKRFLVDLNGQYRIILVK 387

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB ID	Description
1	1938	100.0	387	9	US-09-895-913A-120
2	616.5	31.8	503	10	Sequence 120, Appli
3	584.5	30.2	447	10	Sequence 8, Appli
4	584.5	30.2	475	10	Sequence 2, Appli
5	584.5	30.2	498	10	Sequence 12, Appli
6	459	23.7	549	9	Sequence 11, Appli
7	386	19.9	464	9	Sequence 190, Appli
8	355.5	18.3	286	10	Sequence 182, Appli
9	340.5	17.6	476	10	Sequence 256, Appli
10	289	14.9	330	10	Sequence 184, Appli
11	281.5	14.5	355	9	Sequence 37, Appli
12	211	10.9	596	9	Sequence 161, Appli
13	211	10.9	729	10	Sequence 26, Appli
14	147.5	7.6	397	9	Sequence 2, Appli
15	139	7.2	3241	10	Sequence 280, Appli
16	127	6.6	2041	10	Sequence 1, Appli
17	126.5	6.5	369	10	Sequence 1397, Appli
18	126	6.5	2037	10	Sequence 2, Appli
19	126	6.5	2037	10	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-895-913A-120
; Sequence 120, Application US/09895913A

; Patent No. US20020160456A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles

; APPLICANT: Tomb, Jean Francois

; APPLICANT: Oomen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20020160456A1el Helicobacter Polypeptides in t

; FILE REFERENCE: 06132/043002

; CURRENT APPLICATION NUMBER: US/09/895,913A

; CURRENT FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 08/881,227

; PRIOR FILING DATE: 1997-06-24

; NUMBER OF SEQ ID NOS: 368

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 120

; LENGTH: 387

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-09-895-913A-120

Query Match 100.0%; Score 1938; DB 9; Length 387;

Best Local Similarity 100.0%; Pred. No. 2.1e-140;

Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIPKMERALGSGVLIISKDGYIVTNHNVHDGADKIKVTIPGSKKEYSATLVGTDSDDL 60

Db 1 MIPKMERALGSGVLIISKDGYIVTNHNVHDGADKIKVTIPGSKKEYSATLVGTDSDDL 60

QY 61 AVIRITKDNLPTRKFSNDSNDISVGLVFAIGNPFGVGSVTQGIYSALNKSIGINSYEN 120

Db 61 AVIRITKDNLPTRKFSNDSNDISVGLVFAIGNPFGVGSVTQGIYSALNKSIGINSYEN 120

QY 121 FIOTDASINPGNSGGALIDSRGLGVINTAITSKGGNHGIGFAIPSNMVKDTVTLQIKT 180

Db 121 FIOTDASINPGNSGGALIDSRGLGVINTAITSKGGNHGIGFAIPSNMVKDTVTLQIKT 180

Sequence 6, Appli
Sequence 6, Appli
Sequence 4, Appli
Sequence 179, Appli
Sequence 6, Appli
Sequence 9, Appli
Sequence 10, Appli
Sequence 304, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 34281, A
Sequence 33, Appli
Sequence 302, Appli
Sequence 5352, Ap
Sequence 12278, A
Sequence 11037, A
Sequence 32, Appli
Sequence 2, Appli
Sequence 59, Appli
Sequence 3, Appli
Sequence 5824, Ap
Sequence 12997, A
Sequence 10779, A
Sequence 11589, A
Sequence 63, Appli

QY 181 GKIERGYLVGLQDLSGLNSYDNKEGAVVISVEKDSAPKAGILVWDLITEVNGKKVK 240
DB 181 GKIERGYLVGLQDLSGLNSYDNKEGAVVISVEKDSAPKAGILVWDLITEVNGKKVK 240
QY 241 NTNELNRLIGSMPLPNORVTLKVRDKKERAFITLTLAERKNPNKKEITISAQNGAQQGLNGL 300
DB 241 NTNELNRLIGSMPLPNORVTLKVRDKKERAFITLTLAERKNPNKKEITISAQNGAQQGLNGL 300
QY 301 QVEDLTQETKRSMRLSDVQGVLSQVNSNPAEQAGFRQGNITIKIEEVEKSVADFNH 360
DB 301 QVEDLTQETKRSMRLSDVQGVLSQVNSNPAEQAGFRQGNITIKIEEVEKSVADFNH 360
QY 361 ALEKYKPKRFLVLDLNOGYRIILVK 387
DB 361 ALEKYKPKRFLVLDLNOGYRIILVK 387

RESULT 2

US-09-752-385-8
; Sequence 8, Application US/09752385
; Patent No. US20020012919A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
; and Methods and Compositions for Diagnosing
; Rochalimaea
; Henselae and Rochalimaea Quintana Infection
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSES: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/752,385
; APPLICATION NUMBER: 09/752,385
; FILING DATE: 29-Dec-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/525,310
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-752-385-8

Query Match 31.8%; Score 616.5; DB 10; Length 503;
Best Local Similarity 39.1%; Pred. No. 1.2e-39;
Matches 147; Conservative 63; Mismatches 145; Indels 21; Gaps 8;

QY 10 ALGSGVVISKDGIVTNNHVI-DGADKIKVTIPGSKNKEYSATLVGTDSDLAIVRIT-K 67
DB 125 AFGSGGFSSDGYIVTNNHVISDGTSYAVLDDGT--ELNAKILGTDPRDLAVLKVNEK 182
QY 68 DNLPTTKFSDSDNSISVGLVFAIGNPFGVGSVTQGVISALNKGIGINSYENFIQTDA 127

DB 183 RKFSYVDFGDDSKLRVGDWVVAIGNPFGVGTAGIVSARGRD-IGTGYYDDFIQIDAA 241
QY 128 INPGNSGGALIDSRGGLVGINTAIISKGTGNHGIGFAIPSNMVKDVTVOLIKTKGTERGY 187
DB 242 VNRGNSGGTFDLNGKVGVTNTAIFSPSGNVCIAFAIPAATANEVVQQLIEKGLVQRGW 301
QY 188 LGVGLQDLSGLNSYDNKE--GAVVISVEKDSAPKAGILVWDLITEVNGKKVKNTEL 245
DB 302 LGVQIQVPTKEISDSIGLKEAKGALITDPLK-GPAAKAGIKAGDVIISVNGEKINDVRDL 360
QY 246 RNLIGSMPLPNORVTLKVRDKKERAFITLTLAERKNPNKKEITISAQNGAQQGLNGLQVEDL 305
DB 361 AKRIANMSPGETVTLGVWKSKEENIKVKLDSMPEDENMKDGSKYSEHGN----- 411
QY 306 TOETKRSMRL-----SDVQGVLSQVNSNPAEQAGFRQGNITIKIEEVEKSVADFNH 361
DB 412 SDTLEDYGLIVAPSDDGLGLVVTVDVDPDSDAADKAGIRGDIIVTNNKSVKVSITDT 471
QY 362 LEKYKPKRFLVLDL 377
DB 472 IKNAQKLGKRAILLQV 487
RESULT 3
US-09-388-089B-2
; Sequence 2, Application US/09388089B
; Patent No. US20020018782A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, W.
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID
; FILE REFERENCE: 7969-083
; CURRENT APPLICATION NUMBER: US/09/388,089B
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Neisseria spp.
US-09-388-089B-2
Query Match 30.2%; Score 584.5; DB 10; Length 447;
Best Local Similarity 37.3%; Pred. No. 2.8e-37;
Matches 143; Conservative 78; Mismatches 143; Indels 19; Gaps 10;
QY 2 IPKERMERA---LGSGVVISKDGIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSES 58
DB 62 IPQEEADGGGLNFGSGFIISKDGYILTNTHTVVTGMSIKVLL-NDKREYTKLIGSDVQS 120
QY 59 DLAVTRI-TKDNLPITKFSDSNDISVGLVFAIGNPFGVGSVTQGVISALNKGIGINS 117
DB 121 DVALLKIDATEELPVVVKIGNPKDLKPGEWVAIGAIFGFDNSVTAG-VSAKGRS-LPNES 178
QY 118 YENFIQTDAISNPGNSGGALIDSRGGLVGINTAIISKGTGNHGIGFAIPSNMVKDVTVOL 177
DB 179 YTFPIOTDVAINPGNSGGPLFNLKQGVVINSQIYSRSGFGMGFISFAIPIDVAMNVAEQL 238
QY 178 IKTKIERGYLVGLQDLSGLNSY--DNKEGAVVISVEKDSAPKAGILVWDLITEV 235
DB 239 KNTGKVQQLGVIIQEVSYSLAQSFGLDKAGALIAKILPGSPAERAGLRAGDVLSD 298
QY 236 GKVKVNTNELNRLIGSMPLPNORVTLKVRDKKERAFITLTLAERKNPNKKEITISAQNG 292
DB 299 GGEIRSSGDLPLVWVAIGAITPGKEVSLGVWRKGEITIKVLGNAAEHIGASSKTDEAPYTE 358
QY 293 AQQLNGLQVEDLTQETKRSMRLSDVQGVLSQVNSNPAEQAGFRQGNITIKIEEVEV 352
DB 359 QQSQTFSVESAGITLQTH-----TDSGGHLV-VVRVSDAERAGLRGRLGQVFPV 412
QY 353 KSVADFNNHALEYKKGKPKRFLVL 375

Db 413 NDEAGFRKAMDK-AGKNVPLLM 434

RESULT 4

US-09-388-089B-12
; Sequence 12, Application US/09388089B
; Patent No. US20020018782A1

; GENERAL INFORMATION:

; APPLICANT: Jackson, W.

; APPLICANT: Harris, A.

; TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID

; FILE OF INVENTION: SEQUENCE AND USES THEREOF

; CURRENT APPLICATION NUMBER: 7969-083

; CURRENT FILING DATE: 1999-08-31

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 12

; LENGTH: 475

; TYPE: PRT

; ORGANISM: Neisseria meningitidis

US-09-388-089B-12

Query Match 30.2%; Score 584.5; DB 10; Length 475;
Best Local Similarity 37.3%; Pred. No. 3e-37;
Matches 143; Conservative 78; Mismatches 143; Indels 19; Gaps 10;

QY 2 IPKERMERA---LGSVLIISKDGYIVTNNHVIDGADKTKVTIPGNSKEYSATLVGTDSES 58

Db 90 IQEEADGGGLNFGSGFIISKDGYILTNTHVVTGMSIKVLL-NDKREYTKLIGSDVQS 148

QY 59 DLAVIRI-TKDNLPKIFKSDSDISVGLVFAIGNPFGVGSVTOGIYSALNKSIGINS 117

Db 149 DVALIKIDATELPVVKIGNPKDLKPGEWAAIGAPFDNSVTAG-VSAKGRS-LPNES 206

QY 118 YENFIOTDASINPGNSGGALIDSRGLVGINFAITSKGGNHGIGFAIPSNMVKDTVTQL 177

Db 207 YTPFIOTDVAINPNSGGPLFLNKGVVGINSIYRSRGFGMISFAIPIDVAMNVAEQL 266

QY 178 IKTGKIERGYLGVLQDLSGDLQNSY--DNKEGAVVISVEKDSPPAKKAGILVMDLITEVN 235

Db 267 KNTGVQVRQGLVITQTH-----TDSSGHLV-VVVRSDAERAGLRAGDILVSLD 326

QY 236 GKVKNTNELNLIGSMPLNQRTLVIRDKKERAFITL---ABRKNPKKETISAQNG 292

Db 327 GGEIRSSGDLPMVGAITPGKEVSLGVWRKGEITIKVKGNAEHIGASSKTDEAPYTE 386

QY 293 AQQLNGLOVEDLTQETKRSMLSDVQGLVLSQVNSPAPQAGFROGNIITKIEVEV 352

Db 387 QQSGTFSVESAGITLQTH-----TDSSGHLV-VVVRSDAERAGLRAGDILVSLD 440

QY 353 KSVADFNHALEYKPKRFLVL 375

Db 441 NDEAGFRKAMDK-AGKNVPLLM 462

RESULT 5

US-09-388-089B-11

; Sequence 11, Application US/09388089B
; Patent No. US20020018782A1

; GENERAL INFORMATION:

; APPLICANT: Jackson, W.

; APPLICANT: Harris, A.

; TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID

; FILE OF INVENTION: SEQUENCE AND USES THEREOF

; CURRENT APPLICATION NUMBER: 7969-083

; CURRENT FILING DATE: 1999-08-31

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 11

; LENGTH: 498

; TYPE: PRT

; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-388-089B-11

Query Match 30.2%; Score 584.5; DB 10; Length 498;
Best Local Similarity 37.3%; Pred. No. 3.2e-37;
Matches 143; Conservative 78; Mismatches 143; Indels 19; Gaps 10;

QY 2 IPKERMERA---LGSVLIISKDGYIVTNNHVIDGADKTKVTIPGNSKEYSATLVGTDSES 58

Db 113 IQEEADGGGLNFGSGFIISKDGYILTNTHVVTGMSIKVLL-NDKREYTKLIGSDVQS 171

QY 59 DLAVIRI-TKDNLPKIFKSDSDISVGLVFAIGNPFGVGSVTOGIYSALNKSIGINS 117

Db 172 DVALIKIDATELPVVKIGNPKDLKPGEWAAIGAPFDNSVTAG-VSAKGRS-LPNES 229

QY 118 YENFIOTDASINPGNSGGALIDSRGLVGINFAITSKGGNHGIGFAIPSNMVKDTVTQL 177

Db 230 YTPFIOTDVAINPNSGGPLFLNKGVVGINSIYRSRGFGMISFAIPIDVAMNVAEQL 289

QY 178 IKTGKIERGYLGVLQDLSGDLQNSY--DNKEGAVVISVEKDSPPAKKAGILVMDLITEVN 235

Db 290 KNTGVQVRQGLVITQTH-----TDSSGHLV-VVVRSDAERAGLRAGDILVSLD 349

QY 236 GKVKNTNELNLIGSMPLNQRTLVIRDKKERAFITL---ABRKNPKKETISAQNG 292

Db 350 GGEIRSSGDLPMVGAITPGKEVSLGVWRKGEITIKVKGNAEHIGASSKTDEAPYTE 409

QY 293 AQQLNGLOVEDLTQETKRSMLSDVQGLVLSQVNSPAPQAGFROGNIITKIEVEV 352

Db 410 QQSGTFSVESAGITLQTH-----TDSSGHLV-VVVRSDAERAGLRAGDILVSLD 463

QY 353 KSVADFNHALEYKPKRFLVL 375

Db 464 NDEAGFRKAMDK-AGKNVPLLM 485

RESULT 6

US-09-712-363-190

; Sequence 190, Application US/09712363
; Patent No. US20020164588A1

; GENERAL INFORMATION:

; APPLICANT: Eisenberg, David

; APPLICANT: Roitstein, Sergio H.

; APPLICANT: Marcotte, Edward M.

; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND

; FILE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS

; FILE REFERENCE: 07419-032001

; CURRENT APPLICATION NUMBER: US/09712,363

; CURRENT FILING DATE: 2000-11-13

; PRIOR APPLICATION NUMBER: PCT/US00/02246

; PRIOR FILING DATE: 2000-01-28

; PRIOR APPLICATION NUMBER: 60/179,531

; PRIOR FILING DATE: 2000-02-01

; PRIOR APPLICATION NUMBER: 60/117,844

; PRIOR FILING DATE: 1999-01-29

; PRIOR APPLICATION NUMBER: 60/118,206,

; PRIOR FILING DATE: 1999-02-01

; PRIOR APPLICATION NUMBER: 60/126,593

; PRIOR FILING DATE: 1999-03-26

; PRIOR APPLICATION NUMBER: 60/134,093

; PRIOR FILING DATE: 1999-05-14

; PRIOR APPLICATION NUMBER: 60/134,092

; PRIOR FILING DATE: 1999-05-14

; PRIOR APPLICATION NUMBER: 60/165,124

; PRIOR FILING DATE: 1999-11-12

; PRIOR APPLICATION NUMBER: 60/165,086

; PRIOR FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 292

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 190

; LENGTH: 549

; TYPE: PRT

```
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-190

Query Match      23.7%; Score 459; DB 9; Length 549;
Best Local Similarity 39.9%; Pred No. 1.3e-27;
Matches 113; Conservative 49; Mismatches 99; Indels 22; Gaps 6;

QY 12 GSGVIISKDGYIVTNHVIDG----DKIKVTIP-GSNKEYSATLVGTDSESLAVIRIT 66
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 275 GSGVIVDGRGIVYTNHVIDG----DKIKVTIP-GSNKEYSATLVGTDSESLAVIRIT 66
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

QY 67 K-DNLTPIKFSNDISVGLVFAIGNPFGVSGESVTQGIYSALNK-----SGIGINSYE 119
      |||: | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 335 NVDNLTVARLGDSSKVRVGVDEVAVGAPLGRSTVTQGIYSALHRRPVLSGEGSDTDTVI 394
      |||: | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 120 NFQTODASINPGSGGALIDSRGLVGINTAIISKTGNGHGTGFAIPSNMVKDVTYQLIK 179
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 395 DAIQTDSINHGSGGLIDMDAQVIGITAGKSLSDSASGLGFAIPVNMKLVANSLLIK 454
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

QY 180 TGKIERGYLGVLQDLSGLDLSYDNKSGAVISVEKDSPAKKAGILVWDLITEVNGKKV 239
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 455 DGKIVHTLIGISTRYSNAIAS-----GAQVANVKAGSPAQKGILENDVIVKVGNAV 508
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

QY 240 KTNELNLIGSMLPNQRTVLKVRDKKERAFTLLAERKNPN 282
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 509 ADSDEFVAVRQLAQGDAPIEVREGRH----VTLTVPDPD 547
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

RESULT 7
US-09-712-363-182
; Sequence 182, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-182

Query Match      19.9%; Score 386; DB 9; Length 464;
Best Local Similarity 38.0%; Pred. No. 3.8e-22;
Matches 103; Conservative 42; Mismatches 96; Indels 30; Gaps 8;

QY 12 GSGVIISKDGYIVTNHVIDGADK-----IKVTIPGSN-KEYSATLVGTDSESLAVI 63
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
```

Db 240 IIVKNGRPLVDSSELQE---AVLTESPLLEVRNGNDLLFSI 280

RESULT 9

US-09-935-390A-37
 ; Sequence 37, Application US/09935390A
 ; Patent No. US200200761A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Escobedo, Jaime
 ; Quianjin, Hu
 ; Garcia, Pablo
 ; Williams, Lewis T.
 ; Kothakota, Srinivas
 ; TITLE OF INVENTION: Secreted Human Proteins
 ; NUMBER OF SEQUENCES: 38
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/935,390A
 ; FILING DATE: 22-Aug-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/988,671
 ; FILING DATE: 1997-12-11
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jane E. R. Potter
 ; REGISTRATION NUMBER: 33,332
 ; REFERENCE/DOCKET NUMBER: 1369,002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (510) 923-2718
 ; TELEFAX: (510) 655-3542
 ; TELEX: <Unknown>
 ; INFORMATION FOR SEQ ID NO: 37:
 ; MOLECULE TYPE: No. US200200761A1e
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
 US-09-935-390A-37

Query Match 17.68; Score 340.5; DB 10; Length 476;
 Best Local Similarity 31.94; Pred. No. 1.1e-18;
 Matches 89; Conservative 61; Mismatches 102; Indels 27; Gaps 10;
 Qy 12 GSGVIISKDGIYVNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRIKD-NL 70
 Db 202 GSGFVISEGLIITNAHVVRNQWIEVLIQ-NGARYEAVVXIDILKDLAVIKIESNAEL 260
 Qy 71 PTIKFSDSDNISVGLVFAIGNPFGVGSVTGGIVSALNKSQ--IGI-NSYENFIQTDS 127
 Db 261 PVLMLGRSSDLRAGEFVVALGSPFSLQNTATAGIVSTKQRGKELGKMDSDMDYQIDAT 320
 Qy 128 INPNSGSGALLDSRGLGVINTAISKYGGHNGIGFAIPSNMVKDTVTQLIK---TGKI- 183
 Db 321 INYGNSSGFLVLDGDIVGVNSLRVT-----DGISFAIPSDRVQRQFAEYHEHQMGKAF 375
 Qy 184 -ERGVGLVGLQD---LSGDQLNSY---DNKEGAVVISVEKDSPPAKKAGILVMDLITEV 234
 Db 376 SNKKYLGQLMSLTVPLSEELKMHYPDPDVSSGVYVCKVVEGTAAGSSGLRDHDHVI 435
 Qy 235 NGKKVKNTELNLRLITGMLPNQVRTLKVIRDKKRAFTL 273

Db 436 NGKPITTTTD-----VVKALDSLSMAVLRGKDNLLTV 470

RESULT 10

US-09-764-898-184
 ; Sequence 184, Application US/09764898
 ; Patent No. US20020090673A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PUZ01
 ; CURRENT APPLICATION NUMBER: US/09/764,898
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 311
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 184
 ; LENGTH: 330
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (244)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (265)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-898-184

Query Match 14.98; Score 289; DB 10; Length 330;
 Best Local Similarity 40.08; Pred. No. 5.7e-15;
 Matches 70; Conservative 33; Mismatches 56; Indels 16; Gaps 6;
 Qy 12 GSGVIISKDGIYVNNHVID-----GADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI 65
 Db 148 GSGFIMSEAGLIITNAHVVSNSAARGQOLKVLQNGD-SYEATIKDIDKSDIATIKI 206
 Qy 66 -TKDNLPTIKFSDNISVGLVFAIGNPFGVGSVTGGIVSALNKSQ--IGI-NSYENF 121
 Db 207 HPKKLPVLLGHSAIDLRPFVVAIGSPFALQNTVTGIVSTAQREGRLGRSDDMXY 266
 Qy 122 IQTDASINPNSGSGALLDSRGLGVINTAISKYGGHNGIGFAIPSNMVKDTVTQ 176
 Db 267 IQTDAIINYGNSSGFLVLDGDIVGVNSLRVT-----GISFAIPSDRITRFLTE 316

RESULT 11

US-09-712-363-161
 ; Sequence 161, Application US/09712363
 ; Patent No. US20020164588A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eisenberg, David
 ; APPLICANT: Rotstein, Sergio H.
 ; APPLICANT: Marcotte, Edward M.
 ; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
 ; FILE REFERENCE: 07419-032001
 ; CURRENT APPLICATION NUMBER: US/09/712,363
 ; CURRENT FILING DATE: 2000-11-13
 ; PRIOR APPLICATION NUMBER: PCT/US00/02246
 ; PRIOR FILING DATE: 2000-01-28
 ; PRIOR APPLICATION NUMBER: 60/179,531
 ; PRIOR FILING DATE: 2000-02-01
 ; PRIOR APPLICATION NUMBER: 60/117,844
 ; PRIOR FILING DATE: 1999-01-29
 ; PRIOR APPLICATION NUMBER: 60/118,206,
 ; PRIOR FILING DATE: 1999-02-01
 ; PRIOR APPLICATION NUMBER: 60/126,593
 ; PRIOR FILING DATE: 1999-03-26
 ; PRIOR APPLICATION NUMBER: 60/134,093
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/134,092

```

;
; TYPE: PRT
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Description of Artificial Sequence:bl-fusion
US-09-287-849-26

```

RESULT 14
US-09-712-363-280

```
; Sequence 280, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09712,363
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 280
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-280

Query Match          7.6%; Score 147.5; DB 9; Length 397;
Best Local Similarity 27.1%; Pred. No. 0.00043;
Matches 52; Conservative 33; Mismatches 80; Indels 27; Gaps 8;

Qy 6 RMERAL-CSGVIISKDGIYVNNHVIDGADKIKVTIPGSKKEYSATLVGTDSSEDLAVIR 64
Db 213 RCQKLEGTGTVISPD-RVMTNAHVAGSN--NVTYAGDRPFATVVSYPDSVDVAILA 269

Qy 65 ITKDNLPITKESDSDISVGLVFAIGNPFGVGESVTGQ-IVSALNKGSGIGINSYENFIQ 123
Db 270 VPHLPPPLVFA-REPATGADVVLGVPGGNTATPARIREAIRLSPDIYGDPEPVT 328

Qy 124 TD-----ASINPGSGGALIDSRGLVGI-----NTAISKGTGGNHGIGFAIPSNM 169
Db 329 RDVYTRADVEGDSGGPLIDINGVLGVFGAIDDAETGTVLTAG-----EVAGQL 381

Qy 170 VKDVTQLIKTG 181
Db 382 AKIGATQPVGTG 393

RESULT 15
US-09-841-786-1
; Sequence 1, Application US/09841786
; Patent No. US20020054883A1
; GENERAL INFORMATION:
; APPLICANT: NAGARAJA, T. G.
; APPLICANT: STEWART, GEORGE C.
; APPLICANT: NARAYANAN, SANJEEV K.
; APPLICANT: CHENGAPPA, M. M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN
; TITLE OF INVENTION: VACCINE AND PREPARATION THEREOF
; FILE REFERENCE: 30296
; CURRENT APPLICATION NUMBER: US/09/841,786
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 09/558,257
```

```
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3241
; TYPE: PRT
; ORGANISM: Fusobacterium necrophorum
US-09-841-786-1

Query Match          7.2%; Score 139; DB 10; Length 3241;
Best Local Similarity 19.9%; Pred. No. 0.034;
Matches 99; Conservative 68; Mismatches 156; Indels 174; Gaps 22;

Qy 4 KERMERALGSGVIISKOGIYVNNHV-IDGAD-----KIKVTIPGSKNE- 46
Db 1501 KEQLEKAK-----KKEGAVIVNAALSVAAGTDKSAGGVAIAVNTVKRKFKAELSGSNEA 1554

Qy 47 ----YSATLVGTDSSEDLAVIR-----ITKD-----NLPTIKFSDSDISV 83
Db 1555 GEDKIHAKHVNEAKSSTVVVNAASGLAISKDAFGSMGSGAWQDLSNDTIAKVDKGRISA 1614

Qy 84 GDLVFAIGNPFGVGESVTGQIVSALNKGSGIG-----INSYENFIOTDASINPGNSGGA 136
Db 1615 DSLNVNANNSI-IGVNVAGTTAGLSL-TAVGAAFANNTLHNKTSALITGTVKNPFGSKNT 1672

Qy 137 LID-----SRGGLVGIINTAIIISKTGGNHGIGFAIPSNMVKDVTQLIKTGKIE 184
Db 1673 KVNQALNDSHTVNVSAGGAASIKQA-----GIGMWVSVNRKGSDETEALVSDSEFE 1723

Qy 185 RGYLGVGLQDLSDGLQNSYDNKKEGAVISVEKDSAPKAGILVWDLITVEVNGKKVKNTNE 244
Db 1724 ----GVSSFNVDAKDQKTINTIAG-----NANGGKAAGV-----GATVAHTNI 1762

Qy 245 LRNLIGSMLPNQRYTLKVIROKKE-----RAFTLT----- 274
Db 1763 GKQSVIAIVKNSKITTANDQDRKNINVTAKDYTMTNTTIAVGGGAKGASVOGASASTLN 1822

Qy 275 -----LAERKNPNKKE---TISAQN-----GAQGOL---NGLIQ 301
Db 1823 KTVSSHVDQTDIDKDLSEENNGNKEKANVNVLAENTSQVVTNATVLSGASGOAAVGAVA 1882

Qy 302 VEDLTQETKRSMILS-DDVQGVLVSVQVNS-----PAEQAGFRQGNITTKIEEV 350
Db 1883 VNKITQNTSAHIKUNSTQVNRNALVKSKSHSSTIKTIGICAGVGAGGAGVTVGSVAVNKIYN- 1941

Qy 351 EVKSVADFNHALEKYYG 367
Db 1942 --NTIAELNHAKYITAKG 1956

Search completed: November 18, 2002, 11:33:41
Job time : 14 secs
```



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:28:15 ; Search time 18 Seconds
(without alignments)
632.593 Million cell updates/sec

Title: US-09-895-913A-120
Perfect score: 1938
Sequence: 1 MIPKRMERLALSGVSIISKD.....KPKRFLVLDLNGQYRIILVK 387

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCFUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	725.5	37.4	472	1 US-08-278-091-5	Sequence 5, Appli
2	725.5	37.4	472	1 US-08-483-859-5	Sequence 5, Appli
3	725.5	37.4	472	1 US-08-472-173-5	Sequence 5, Appli
4	725.5	37.4	472	2 US-08-487-167-5	Sequence 5, Appli
5	725.5	37.4	472	2 US-08-482-816-5	Sequence 5, Appli
6	725.5	37.4	472	2 US-08-296-149-5	Sequence 5, Appli
7	725.5	37.4	472	2 US-08-801-499-5	Sequence 5, Appli
8	725.5	37.4	472	2 US-08-615-271-5	Sequence 5, Appli
9	725.5	37.4	472	3 US-09-074-660-5	Sequence 5, Appli
10	725.5	37.4	472	3 US-09-074-659-5	Sequence 5, Appli
11	725.5	37.4	472	3 US-09-106-468-5	Sequence 5, Appli
12	725.5	37.4	472	4 US-09-106-466A-5	Sequence 5, Appli
13	725.5	37.4	472	4 US-09-106-467-5	Sequence 5, Appli
14	724.5	37.4	475	1 US-08-278-091-6	Sequence 6, Appli
15	724.5	37.4	475	1 US-08-483-859-6	Sequence 6, Appli
16	724.5	37.4	475	1 US-08-472-173-6	Sequence 6, Appli
17	724.5	37.4	475	2 US-08-487-167-6	Sequence 6, Appli
18	724.5	37.4	475	2 US-08-482-816-6	Sequence 6, Appli
19	724.5	37.4	475	2 US-08-296-149-6	Sequence 6, Appli
20	724.5	37.4	475	2 US-08-801-499-6	Sequence 6, Appli
21	724.5	37.4	475	2 US-08-615-271-6	Sequence 6, Appli
22	724.5	37.4	475	3 US-09-074-660-6	Sequence 6, Appli
23	724.5	37.4	475	3 US-09-074-659-6	Sequence 6, Appli
24	724.5	37.4	475	3 US-09-106-468-6	Sequence 6, Appli
25	724.5	37.4	475	4 US-09-106-466A-6	Sequence 6, Appli
26	724.5	37.4	475	4 US-09-106-467-6	Sequence 6, Appli
27	720.5	37.2	491	3 US-08-923-454A-14	Sequence 14, Appli

28	714.5	36.9	475	1	US-08-350-741-2	Sequence 2, Appli
29	714.5	36.9	475	2	US-08-463-875A-2	Sequence 2, Appli
30	714.5	36.9	492	1	US-08-350-741-3	Sequence 3, Appli
31	714.5	36.9	492	2	US-08-463-875A-3	Sequence 3, Appli
32	706.5	36.5	460	4	US-09-199-637A-132	Sequence 132, App
33	684	35.3	463	1	US-08-485-569-2	Sequence 2, Appli
34	684	35.3	463	1	US-08-480-993-2	Sequence 2, Appli
35	684	35.3	463	2	US-07-903-079B-2	Sequence 2, Appli
36	666	34.4	463	1	US-08-278-091-2	Sequence 2, Appli
37	666	34.4	463	1	US-08-483-859-2	Sequence 2, Appli
38	666	34.4	463	1	US-08-472-173-2	Sequence 2, Appli
39	666	34.4	463	2	US-08-487-167-2	Sequence 2, Appli
40	666	34.4	463	2	US-08-482-816-2	Sequence 2, Appli
41	666	34.4	463	2	US-08-296-149-2	Sequence 2, Appli
42	666	34.4	463	2	US-08-801-499-2	Sequence 2, Appli
43	666	34.4	463	2	US-08-615-271-2	Sequence 2, Appli
44	666	34.4	463	3	US-09-074-660-2	Sequence 2, Appli
45	666	34.4	463	3	US-09-074-659-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-278-091-5
; Sequence 5, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSNORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278.091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-371
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 595-1155
; TELEFAX: (415) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-278-091-5

Query Match 37.4%; Score 725.5; DB 1; Length 472;
Best Local Similarity 43.28; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;
Oy 10 ALGSGVSIISKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSGLAVIRI-TK 67
||||||| | ||:|||||:| ||| : :|| | ||:|:|

```
Db 112 ALGSGVVIDADKGYVYVNNHVDNATVIKVL--SDGRKFDKAMVGDPRSDIALIOIQNP 170
QY 68 DNLPTIKFSDSDISVGLVFAIGNPFGVGSVTQIGVSALNKSIGIGINSYENFIQTDA 127
Db 171 KNLTAIKMADSDALRVGDTYVIGNPFGLGETVTSIGVSALGRSGLNAENYENFIQTDA 230
QY 128 INPNSGGALIDSRGLVGINTAIISKTGNGHIGIFAIPSNMVKDVTVTLIKTKIERGY 187
Db 231 INRNSGGALVNLNGELIGINTAILAPDGGNIGIGIFAIPSNMVKNLTSMQVEYGVKRG 290
QY 188 LGVGLQDLSDGLQNS--YDNKEGAVVISVEKDSPAKAGILVMDLITEVNGKVKVKNEL 245
Db 291 LGIMGTLSNELAKAMKVDAQRGAQFVSQVLPNSSAAKAGIKAGDVITSLNGKPISSFA 350
QY 246 RNLIGSMLPNQRVTLKVIDRKKERAFTLTIAERKNPNKKTETISAQNGAQLNGLQV 305
Db 351 RAQVGTMPVGSKITLGLLRDQGVNVL--QOSSNQVDSSTFNGIEG----- 399
QY 306 TQETKSRMLSDVQGVLYSOVNSPBAQAFRQGNITKIEEVEKVSADFNHALEKY 365
Db 400 ---AEMSNKGD--QGVVYNNVKTGTPTAAQIGLKGVDVIIGANQIAVKNAIEIRKVL 452
QY 366 KGKPKRFLVDLNOGYR 382
453 -SKPS-VLALNTQRGDR 467

RESULT 2
US-08-483-859-5
; Sequence 5, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; REDUCED PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,859
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-483-859-5
Query Match 37.4%; Score 725.5; DB 1; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;
QY 10 ALGSGVVIISKD--GYIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSSEDLAVIRI-TK 67
Db 112 ALGSGVVIDADKGYVYVNNHVDNATVIKVL--SDGRKFDKAMVGDPRSDIALIOIQNP 170
QY 68 DNLPTIKFSDSDISVGLVFAIGNPFGVGSVTQIGVSALNKSIGIGINSYENFIQTDA 127
Db 171 KNLTAIKMADSDALRVGDTYVIGNPFGLGETVTSIGVSALGRSGLNAENYENFIQTDA 230
QY 128 INPNSGGALIDSRGLVGINTAIISKTGNGHIGIFAIPSNMVKDVTVTLIKTKIERGY 187
Db 231 INRNSGGALVNLNGELIGINTAILAPDGGNIGIGIFAIPSNMVKNLTSMQVEYGVKRG 290
QY 188 LGVGLQDLSDGLQNS--YDNKEGAVVISVEKDSPAKAGILVMDLITEVNGKVKVKNEL 245
Db 291 LGIMGTLSNELAKAMKVDAQRGAQFVSQVLPNSSAAKAGIKAGDVITSLNGKPISSFA 350
QY 246 RNLIGSMLPNQRVTLKVIDRKKERAFTLTIAERKNPNKKTETISAQNGAQLNGLQV 305
Db 351 RAQVGTMPVGSKITLGLLRDQGVNVL--QOSSNQVDSSTFNGIEG----- 399
QY 306 TQETKSRMLSDVQGVLYSOVNSPBAQAFRQGNITKIEEVEKVSADFNHALEKY 365
Db 400 ---AEMSNKGD--QGVVYNNVKTGTPTAAQIGLKGVDVIIGANQIAVKNAIEIRKVL 452
QY 366 KGKPKRFLVDLNOGYR 382
Db 453 -SKPS-VLALNTQRGDR 467

RESULT 3
US-08-472-173-5
; Sequence 5, Application US/08472173
; Patent No. 5655353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; REDUCED PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,173
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
```

; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-173-5

Query Match 37.4%; Score 725.5; DB 1; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;
Qy 10 ALGSGVVISKD-GYIVTNHVIDGADTKVTPGSKNKEYSATLVGTDSSEDLAVIRI-TK 67
Db 112 ALGSGVVIDADKGVVTNNHVVNDATVIKVL-SDGRKFDKAMVKGKPRSDIALIQINP 170
Qy 68 DNLPTIKFSDNDISVGLVFAIGNPGVGSVTVQGIIVSALNKGSGINGSNENFIQTDA 127
Db 171 KNLTAIRKWDSDALRVGDYTVGIGNPGLGETVTSVIGVLSALGRSGLNAENYFIQTDA 230
Qy 128 INPGSGALIDSRGLVGINTAISKGTGNHIGFPAIPSNMVKDVTVQLIKTKIERGY 187
Db 231 INRSGGALVNLGELIGINTAILAPDGGNIGIGFPAIPSNMVKNTSQWVEYGVQRKE 290
Qy 188 LGVGLQDLSGLQNS--YDNKEGAVISVEKDSAPKAGILVLDLITEVNGKKVKNTEL 245
Db 291 LGIMGTLSNLSLAKAMKVDAQRGAQVSPVLPNSSAAKAGIKAGDVITSLNGKPISSFAAL 350
Qy 246 RNLIGSMPLPNQVTLKVIKDKKERAFTLTLAERNPNKKTETISAQNGAOGGLNGLOVEDL 305
Db 351 RAQVGTMPVGSKLTGLLRDQKQVNVNLEL--QQSQSQVDSSTFNGIEG----- 399
Qy 306 TQETKSRMRLSDDYQGVVLSQVNSPNAEQAGFRQGNITIKIEEVEKVSADFNHALEKY 365
Db 400 --AEMSNKGD--QGVVNVNKTGTTPAAQIGLKKGDVIGANQIAVKNIAEIRKVL-- 452
Qy 366 KGKPKRFLVLDLNOGYR 382
b 453 -SKPS-VLALNIQRGR 467

RESULT 4

US-08-472-167-5
; Sequence 5, Application US/08487167
; Patent No. 5869302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-487-167-5

Query Match 37.4%; Score 725.5; DB 2; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;
Qy 10 ALGSGVVISKD-GYIVTNHVIDGADTKVTPGSKNKEYSATLVGTDSSEDLAVIRI-TK 67
Db 112 ALGSGVVIDADKGVVTNNHVVNDATVIKVL-SDGRKFDKAMVKGKPRSDIALIQINP 170
Qy 68 DNLPTIKFSDNDISVGLVFAIGNPGVGSVTVQGIIVSALNKGSGINGSNENFIQTDA 127
Db 171 KNLTAIRKWDSDALRVGDYTVGIGNPGLGETVTSVIGVLSALGRSGLNAENYFIQTDA 230
Qy 128 INPGSGALIDSRGLVGINTAISKGTGNHIGFPAIPSNMVKDVTVQLIKTKIERGY 187
Db 231 INRSGGALVNLGELIGINTAILAPDGGNIGIGFPAIPSNMVKNTSQWVEYGVQRKE 290
Qy 188 LGVGLQDLSGLQNS--YDNKEGAVISVEKDSAPKAGILVLDLITEVNGKKVKNTEL 245
Db 291 LGIMGTLSNLSLAKAMKVDAQRGAQVSPVLPNSSAAKAGIKAGDVITSLNGKPISSFAAL 350
Qy 246 RNLIGSMPLPNQVTLKVIKDKKERAFTLTLAERNPNKKTETISAQNGAOGGLNGLOVEDL 305
Db 351 RAQVGTMPVGSKLTGLLRDQKQVNVNLEL--QQSQSQVDSSTFNGIEG----- 399
Qy 306 TQETKSRMRLSDDYQGVVLSQVNSPNAEQAGFRQGNITIKIEEVEKVSADFNHALEKY 365
Db 400 --AEMSNKGD--QGVVNVNKTGTTPAAQIGLKKGDVIGANQIAVKNIAEIRKVL-- 452
Qy 366 KGKPKRFLVLDLNOGYR 382
Db 453 -SKPS-VLALNIQRGR 467

RESULT 5

US-08-482-816-5
; Sequence 5, Application US/08482816
; Patent No. 5935573
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease A
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto

STATE: Ontario
COUNTRY: Canada
ZIP: MSG 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,816
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-482-816-5

Query Match 37.4%; Score 725.5; DB 2; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;

Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVVISKD-GYIVTNHHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
DB 112 ALGSGVIIDADKGYVTVNNHVVVDNATVIKVL-SDGRKFDKMKVGDPRSDIALIQINP 170
QY 68 DNLPTIKFSDSDISVGLVFAIGNPFGVGSVTOGIVSALNKGIGINSYENFIQTDA 127
DB 171 KNLTAIKMADSDALRVGDTYVIGNPFGLGETVTSIGVSALGRSLNAENYENFIQTDA 230
QY 128 INPNSGGLIDSRGLVGINTAIISKGTGNHIGIGFAIPSNMVKDVTOLIKTKIERGY 187
DB 231 INRNSGGALVNLGELIGINTAILAPDGNIGIGFAIPSNMVKNTLSQMEYGVKRG 290
QY 188 LGVGLQDLSGDLQNS--YDNKEGAVVISVEKDSPAKKAGILVMDLITEVNGKVKNTNEL 245
DB 291 LGIMGTLSNELAKAMKVDAQGFVSVQLPNSSAAKAGIKAGDVITSLNGKPISSFAAL 350
QY 246 RNLGSMPLNQRTVLKVRDKKERAFTLTLAERKPNKKTETISAONGAQGLNGLQVEDL 305
DB 351 RAQGTMPVSKLTGLLRDGRQVNVNLEL-QQSSQNQVDDSSIFNGIEG----- 399
QY 306 TOETKRMRSLDDVQGLVSVQNSNPAPQAGFROGNIITKTEEVKSVADFNHALEKY 365
DB 400 ---AEMSNKGD--QGVVNVNKTGTPTAAQIGLKKGDVIGANQIAVKNIAIRKVLDD-- 452

RESULT 6

US-08-296-149-5
; Sequence 5, Application US/08296149
; Patent No. 5939297
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping

APPLICANT: CHONG, Pele
APPLICANT: COHEN, Raymond P.
APPLICANT: KLEIN, Michael H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
REDUCED PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: MSG 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,149
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-390
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-296-149-5

Query Match 37.4%; Score 725.5; DB 2; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;

Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVVISKD-GYIVTNHHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
DB 112 ALGSGVIIDADKGYVTVNNHVVVDNATVIKVL-SDGRKFDKMKVGDPRSDIALIQINP 170
QY 68 DNLPTIKFSDSDISVGLVFAIGNPFGVGSVTOGIVSALNKGIGINSYENFIQTDA 127
DB 171 KNLTAIKMADSDALRVGDTYVIGNPFGLGETVTSIGVSALGRSLNAENYENFIQTDA 230
QY 128 INPNSGGLIDSRGLVGINTAIISKGTGNHIGIGFAIPSNMVKDVTOLIKTKIERGY 187
DB 231 INRNSGGALVNLGELIGINTAILAPDGNIGIGFAIPSNMVKNTLSQMEYGVKRG 290
QY 188 LGVGLQDLSGDLQNS--YDNKEGAVVISVEKDSPAKKAGILVMDLITEVNGKVKNTNEL 245
DB 291 LGIMGTLSNELAKAMKVDAQGFVSVQLPNSSAAKAGIKAGDVITSLNGKPISSFAAL 350
QY 246 RNLGSMPLNQRTVLKVRDKKERAFTLTLAERKPNKKTETISAONGAQGLNGLQVEDL 305
DB 351 RAQGTMPVSKLTGLLRDGRQVNVNLEL-QQSSQNQVDDSSIFNGIEG----- 399
QY 306 TOETKRMRSLDDVQGLVSVQNSNPAPQAGFROGNIITKTEEVKSVADFNHALEKY 365
DB 400 ---AEMSNKGD--QGVVNVNKTGTPTAAQIGLKKGDVIGANQIAVKNIAIRKVLDD-- 452

RESULT 7

US-08-801-499-5
; Sequence 5, Application US/08801499

```
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,499
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-671 MIS:jbb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-801-499-5

Query Match 37.4%; Score 725.5; DB 2; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVIISKD-GYIVTNNHVDGADKIKVTIPGNSKEYSATLVGTDSESLAVIRI-TK 67
DB 112 ALGSGVIIDAKGYIVVNNHVDNATVTKVQL-SDGRKFDAMVGKDPKPSDIALIQTNP 170
QY 68 DNLPIKFSDSNDISVGLVFAIGNPFGVGSVTOGIYSALNKGSGIGNSYENFIQTDA 127
DB 171 KNLTAIKWADSDALRVGDTYVIGNPFGLGETVTSIGVSGALGRSGLNAYENFIQTDA 230
QY 128 INPGSGGALIDSRGLVGINTALISKTGGNHGIGFAIPSNMVKDVTVQLIKTGKIERGY 187
DB 231 INRNGSGGALVNLGELIGINTAILAPDGNIGIGFAIPSNMVKNLTSQVVEYGVQVKRGE 290
QY 188 LGVGLQDLSGDLQNS--YDNREGAVVISVEKDSPPAKKAGILVWDLITEVNGKKVNTNEL 245
DB 291 LGIMGTSELSELAKMKYDAQRGAFVSQVLPNSSAAKAGIKAGDVITSLNGKPISSFAAL 350
QY 246 RNLIGSMLPNORVTLKVIIRDKERAFVTLTAERKKNPKNETISAQNGAQGLQGLQVDEL 305
DB 351 RAQVGTMPPVSGKLTGLLRDGRQVNVNLEL-QQSSQNQVSSSIFNGIEG----- 399
```

```
QY 306 TQETKRSMRLSIDVQGVLVSOVNENSPAEQAGFRGQNIITKIEEVEYKSVADFNALEKY 365
DB 400 ---AEMSNKGKI--GGVVNNVKTGTTPAAQIGLKGKGDVIGANQIAVKNIAIRKVL-- 452
QY 366 KGKPKRFLVLDINQGYR 382
DB 453 -SKPS-VLALNIQRGDR 467

RESULT 8
US-08-615-271-5
; Sequence 5, Application US/08615271
; Patent No. 5981503
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-615-271-5
```

```
Query Match 37.4%; Score 725.5; DB 2; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVIISKD-GYIVTNNHVDGADKIKVTIPGNSKEYSATLVGTDSESLAVIRI-TK 67
DB 112 ALGSGVIIDAKGYIVVNNHVDNATVTKVQL-SDGRKFDAMVGKDPKPSDIALIQTNP 170
QY 68 DNLPIKFSDSNDISVGLVFAIGNPFGVGSVTOGIYSALNKGSGIGNSYENFIQTDA 127
DB 171 KNLTAIKWADSDALRVGDTYVIGNPFGLGETVTSIGVSGALGRSGLNAYENFIQTDA 230
QY 128 INPGSGGALIDSRGLVGINTALISKTGGNHGIGFAIPSNMVKDVTVQLIKTGKIERGY 187
DB 231 INRNGSGGALVNLGELIGINTAILAPDGNIGIGFAIPSNMVKNLTSQVVEYGVQVKRGE 290
QY 188 LGVGLQDLSGDLQNS--YDNREGAVVISVEKDSPPAKKAGILVWDLITEVNGKKVNTNEL 245
DB 291 LGIMGTSELSELAKMKYDAQRGAFVSQVLPNSSAAKAGIKAGDVITSLNGKPISSFAAL 350
```

QY 246 RNLIGMLPNQVRVTLKVIIRDKKERAFVTLAERKNPNKKTETISAQNGAQQLNGLOVEDL 305
Db 351 RAQVGTMPVGSKLTLGLLRDQGVNNLEL-QOSSONQVDSISIFNGIEG----- 399
QY 306 TQETKSMRLSDVGVLYSOVNNENSPAQAGFRQGNITIKTIEEVEVKSVADEPNHALEY 365
Db 400 ---AEMSNKGD--QGVVNNVKTGTPTAAQIGLKGDDVIIGANQIAVKNIAEIRKVL-- 452
QY 366 KGPKRFLVLDLNOGYR 382
Db 453 -SKPS-VLALNIQGRDR 467

RESULT 9
US-09-074-660-5
; Sequence 5, Application US/09074660
; Patent No. 6020183
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-731 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-074-660-5

Query Match 37.4%; Score 725.5; DB 3; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVLIISKD-GYIVTNNHVDGADKIKVTIPGSNKEVSATLVGTDSGLAVIRI-TK 67
Db 112 ALGSGVLIIDADKGYVVTNNHVDNATVTKVLQ-SDGRKFDKRWGDKPRSDIALIQNP 170

QY 68 DNLPTIKFSDSNDISVGDVLFALGNPFGYGESVTOGIVSALNKSIGIGINSYENFIQTDA 127
Db 171 KNLTAIKMADSDALRVGDTVGVIGNPFGELGETVTSIGIVSALGRSGLNAENYENFIQTDA 230
QY 128 INPGNSGALIDSRGLGVINTAILISKTGTGNHGIGFAIPSNMVKQVTVTLIKTKIERGY 187
Db 231 INRNSGVALNGLINGELIGINTAILAILPDGNGIGIGFAIPSNMVKNTLSQMVYGVYKRG 290
QY 188 LGVGLQDLSLDLQNS--YDNKEGAVVISVEKQSPAKKAGILVWDLITVEVNGKKVKNTEL 245
Db 291 LGIMGTLSNELAKAMKVDAQRGAFVQVLPNSSAAKAGIKAGDVITSLNGKRFISSFAAL 350
QY 246 RNLIGMLPNQVRVTLKVIIRDKKERAFVTLAERKNPNKKTETISAQNGAQQLNGLOVEDL 305
Db 351 RAQVGTMPVGSKLTLGLLRDQGVNNLEL-QOSSONQVDSISIFNGIEG----- 399
QY 306 TQETKSMRLSDVGVLYSOVNNENSPAQAGFRQGNITIKTIEEVEVKSVADEPNHALEY 365
Db 400 ---AEMSNKGD--QGVVNNVKTGTPTAAQIGLKGDDVIIGANQIAVKNIAEIRKVL-- 452
QY 366 KGPKRFLVLDLNOGYR 382
Db 453 -SKPS-VLALNIQGRDR 467

RESULT 10
US-09-074-659-5
; Sequence 5, Application US/09074659
; Patent No. 6025342
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,659
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-730 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-659-5

Query Match 37.4%; Score 725.5; DB 3; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVILSKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
Db 112 ALGSGVILSKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
QY 68 DNLPTIKESDSNDISVGLVFAIGNPFGVGSVTOGIYSALNKSGLGINSYENFIOTDAS 127
Db 171 KNLTAIKWADSDALRVGDTYVIGNPFGLGEIVTSGIVSALGRSGLNAENYENFIOTDAA 230
QY 128 INPGSGGALIDSRGLVGINTAIISKTGGNHGIGFAIPSNMVKDTVTQIKTKIERGY 187
Db 231 INRSGSGALVNLGELIGINTAILAPDGNIGIGFAIPSNMVKNTLSQMEYGVQVKKRGE 290
QY 188 LGVGLQDLSGLDONS--YDNKEGAVVISVEKDSPAKKAGILVWDLITEVNGKKVNTNEL 245
Db 291 LGIMGTSELNSELAKAMKVDAQRGAFAVSQVLPNNSAAGIKAGIDVITSUNGKPISSFAAL 350
QY 246 RNLGSMPLPNQVTLKIRDKKERAFTLTIAERKNPNKKTETISAQNGAQLNGLOVEDL 305
Db 351 RAQVGTMPVGSKLTLGLLRDQGVNVL--QSSONQVDSISIFNGIEG----- 399
QY 306 TQETKSRMLSDVGVLSQVNSPAPQAGFRQGNITKIEEYKSVADFNHALEKY 365
Db 400 ---AEMSNKGD--QGVVNVNKTGTTPAAQIGLKKGDVIGANQIAVKNIAIRKVLDD-- 452
QY 366 KGKPRFLVLDLNOGYR 382
Db 453 -SKPS-VLALNIQRGRD 467

RESULT 11
US-09-106-468-5
; Sequence 5, Application US/09106468
; Patent No. 6114125
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106.468
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-825
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-106-468-5

Query Match 37.4%; Score 725.5; DB 3; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVILSKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
Db 112 ALGSGVILSKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
QY 68 DNLPTIKESDSNDISVGLVFAIGNPFGVGSVTOGIYSALNKSGLGINSYENFIOTDAS 127
Db 171 KNLTAIKWADSDALRVGDTYVIGNPFGLGEIVTSGIVSALGRSGLNAENYENFIOTDAA 230
QY 128 INPGSGGALIDSRGLVGINTAIISKTGGNHGIGFAIPSNMVKDTVTQIKTKIERGY 187
Db 231 INRSGSGALVNLGELIGINTAILAPDGNIGIGFAIPSNMVKNTLSQMEYGVQVKKRGE 290
QY 188 LGVGLQDLSGLDONS--YDNKEGAVVISVEKDSPAKKAGILVWDLITEVNGKKVNTNEL 245
Db 291 LGIMGTSELNSELAKAMKVDAQRGAFAVSQVLPNNSAAGIKAGIDVITSUNGKPISSFAAL 350
QY 246 RNLGSMPLPNQVTLKIRDKKERAFTLTIAERKNPNKKTETISAQNGAQLNGLOVEDL 305
Db 351 RAQVGTMPVGSKLTLGLLRDQGVNVL--QSSONQVDSISIFNGIEG----- 399
QY 306 TQETKSRMLSDVGVLSQVNSPAPQAGFRQGNITKIEEYKSVADFNHALEKY 365
Db 400 ---AEMSNKGD--QGVVNVNKTGTTPAAQIGLKKGDVIGANQIAVKNIAIRKVLDD-- 452
QY 366 KGKPRFLVLDLNOGYR 382
Db 453 -SKPS-VLALNIQRGRD 467

RESULT 12
US-09-106-466A-5
; Sequence 5, Application US/09106466A
; Patent No. 6147057
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/106,466A
FILING DATE: 514514
CLASSIFICATION: 514514
PRIOR APPLICATION DATA: US 08/615,271
FILING DATE: 20-JUN-1996
CLASSIFICATION: 514514
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-826
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-106-466A-5

Query Match 37.4%; Score 725.5; DB 4; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;
QY 10 ALGSGVLIISKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
Db 112 ALGSGVIIDADRGYVYVNNHVDNATVIKVL-SDGRKFDKMWGKDRSDIALIQNP 170
QY 68 DNLPTIKFSDISNDISVGLVFAIGNPFGVSGESVTOGIVSALNKGIGINSYENFIOTDAS 127
Db 171 KNLTAIKMADSDALRVGDTVGVGNPFGLETGVTSGIVSALGRSGLNAENYENFIOTDAA 230
QY 128 INPGNSGGLIDSRGLVGINTAIISKTGTGNHGIGIPAPSNMVKDVTOLIKTKIERGY 187
Db 231 INNGSGGALVNLGELIGINTAILAPDGGNIGIGIPAPSNMVKNTSQMVEYGQVKRGE 290
QY 188 LGVGLQDLSGDLQNS--YDNKEGAVVISVEKDSPPAKKAGILVWDLITEVNGKVKNTNEL 245
Db 291 LGIMGTLSNLSAKAMKVDQARGAFVSQVLPNNSAAKAGIKAGDVTSLNGKPISSFAAL 350
QY 246 RNLIGSMLPNQRTLVKIRDKKERAFVTLAERKNPNKKTETISAQNGAQQLNGLQVEDL 305
Db 351 RAQVGTMPVGSKLTLGLLRDQGVNVNLEL-QOSSNQVDSSSIFNGIEG-----399
QY 306 TQETKRSMLSDVQGVLYSOVNSPAPQAGFRQGNITTKIEEVEKSVADFNHALEYK 365
Db 400 ---AEMSNKGD--QGVVYNNVKTGTTPAAQIGLKGDDVIIGANQIAVKNAIEIRKVLDD-- 452
QY 366 KGKPRFLVDLNLQGYR 382
Db 453 -SKPS-VLALNIQRGDR 467

RESULT 13
US-09-106-467-5
Sequence 5, Application US/09106467
Patent No. 6153580
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN7 WITH REDUCED
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario

COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,467
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/615,271
FILING DATE: 20-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-824
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-106-467-5

Query Match 37.4%; Score 725.5; DB 4; Length 472;
Best Local Similarity 43.2%; Pred. No. 1.4e-53;
Matches 163; Conservative 75; Mismatches 114; Indels 25; Gaps 9;
QY 10 ALGSGVLIISKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
Db 112 ALGSGVIIDADRGYVYVNNHVDNATVIKVL-SDGRKFDKMWGKDRSDIALIQNP 170
QY 68 DNLPTIKFSDISNDISVGLVFAIGNPFGVSGESVTOGIVSALNKGIGINSYENFIOTDAS 127
Db 171 KNLTAIKMADSDALRVGDTVGVGNPFGLETGVTSGIVSALGRSGLNAENYENFIOTDAA 230
QY 128 INPGNSGGLIDSRGLVGINTAIISKTGTGNHGIGIPAPSNMVKDVTOLIKTKIERGY 187
Db 231 INNGSGGALVNLGELIGINTAILAPDGGNIGIGIPAPSNMVKNTSQMVEYGQVKRGE 290
QY 188 LGVGLQDLSGDLQNS--YDNKEGAVVISVEKDSPPAKKAGILVWDLITEVNGKVKNTNEL 245
Db 291 LGIMGTLSNLSAKAMKVDQARGAFVSQVLPNNSAAKAGIKAGDVTSLNGKPISSFAAL 350
QY 246 RNLIGSMLPNQRTLVKIRDKKERAFVTLAERKNPNKKTETISAQNGAQQLNGLQVEDL 305
Db 351 RAQVGTMPVGSKLTLGLLRDQGVNVNLEL-QOSSNQVDSSSIFNGIEG-----399
QY 306 TQETKRSMLSDVQGVLYSOVNSPAPQAGFRQGNITTKIEEVEKSVADFNHALEYK 365
Db 400 ---AEMSNKGD--QGVVYNNVKTGTTPAAQIGLKGDDVIIGANQIAVKNAIEIRKVLDD-- 452
QY 366 KGKPRFLVDLNLQGYR 382
Db 453 -SKPS-VLALNIQRGDR 467

RESULT 14
US-08-278-091-6
Sequence 6, Application US/08278091
Patent No. 5506139
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.

;; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
;; NUMBER OF SEQUENCES: 23
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sim & McBurney
;; STREET: Suite 701, 330 University Avenue
;; CITY: Toronto
;; STATE: Ontario
;; COUNTRY: Canada
;; ZIP: M5G 1R7

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/278,091
;; FILING DATE: 21-JUL-1994

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Stewart, Michael I

;; REGISTRATION NUMBER: 24,973

;; REFERENCE/DOCKET NUMBER: 1038-371

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (416) 595-1155

;; TELEFAX: (416) 595-1163

;; INFORMATION FOR SEQ ID NO: 6:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 475 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

US-08-278-091-6

Query Match 37.48; Score 724.5; DB 1; Length 475;
Best Local Similarity 43.28; Pred. No. 1.7e-53;
Matches 162; Conservative 74; Mismatches 114; Indels 25; Gaps 9;

Qy 10 ALGSGVIISKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSSESLAVIRI-TK 67
Db 113 ALGSGVIIDADKGYVVTNNHVVDNASVIKVL-SDGRKFDKAVGVGKDRSDIALIQNP 171
Qy 68 DNLPTIKFSDNDSVGLVFAIGNPFGVGSVTQGIYSALNKGSGINSYENFIQTDAAS 127
Db 172 KNLTAIKLADSDALRVGDTVAIGNPFGTGTSGIYVSGALGRSLNENYENFIQTDA 231
Qy 128 INPGSGALIDSRGLVGINTAIISKTGNGHIGIFAIPSNMVKDVTVLTKTGIERGY 187
Db 232 INRNGSGALVNLGELIGINTAILAPDGNIGIGIFAIPSNMVKNLTSQMYEYGOVRGE 291
Qy 188 LGVGLQDLSGLQNS--YDNKEGAVVISVEKDSAPKAGILVWDLITVEVNGKVKNTNEL 245
Db 292 LGIMGTLSNELAKAMKYDAQRGAFVSQVMPNSSAAKAGIKAGDVITSLNGKPISSFAAL 351
Qy 246 RNLIGSMPLPNORVTLKVIIRDKKERAFTLTLAERKNPNKKTISAQNGAQQGLNGQVDEL 305
Db 352 RAQVGTMPVGSKISLGLREGK--AITVNL-----ELQQSSQSQVDSSTIFSG 397
Qy 306 TQETKRSRLSDVGVVLVSQVNSPAPQAGFRGNIITKIEEVEYKSVADFNALEKY 365
Db 398 IEGAEKSNKGD--KGVVVSSVKANSAPAAQIGLKKGDVIGANQIPVKNIAEIRKILD-- 453
Qy 366 KGKPKRFLVLDLNOG 380
Db 454 -SKPS-VLALNIQRG 466

RESULT 15

US-08-483-859-6

;; Sequence 6, Application US/08483859

;; Patent No. 5656436

;; GENERAL INFORMATION:

;; APPLICANT: LOOSMORE, Sheena M.

;; APPLICANT: YANG, Yan-Ping
;; APPLICANT: CHONG, Pele
;; APPLICANT: OOMEN, Raymond P.
;; APPLICANT: KLEIN, Michel H.
;; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
;; TITLE OF INVENTION: Reduced Protease Activity
;; NUMBER OF SEQUENCES: 23
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sim & McBurney
;; STREET: Suite 701, 330 University Avenue
;; CITY: Toronto
;; STATE: Ontario
;; COUNTRY: Canada
;; ZIP: M5G 1R7

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/483,859

;; FILING DATE: 07-JUN-1995

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/296,149

;; FILING DATE: 26-AUG-1994

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/278,091

;; FILING DATE: 21-JUL-1994

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Stewart, Michael I.

;; REGISTRATION NUMBER: 24,973

;; REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (416) 595-1155

;; TELEFAX: (416) 595-1163

;; INFORMATION FOR SEQ ID NO: 6:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 475 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

US-08-483-859-6

Query Match 37.48; Score 724.5; DB 1; Length 475;
Best Local Similarity 43.28; Pred. No. 1.7e-53;
Matches 162; Conservative 74; Mismatches 114; Indels 25; Gaps 9;

Qy 10 ALGSGVIISKD-GYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSSESLAVIRI-TK 67
Db 113 ALGSGVIIDADKGYVVTNNHVVDNASVIKVL-SDGRKFDKAVGVGKDRSDIALIQNP 171
Qy 68 DNLPTIKFSDNDSVGLVFAIGNPFGVGSVTQGIYSALNKGSGINSYENFIQTDAAS 127
Db 172 KNLTAIKLADSDALRVGDTVAIGNPFGTGTSGIYVSGALGRSLNENYENFIQTDA 231
Qy 128 INPGSGALIDSRGLVGINTAIISKTGNGHIGIFAIPSNMVKDVTVLTKTGIERGY 187
Db 232 INRNGSGALVNLGELIGINTAILAPDGNIGIGIFAIPSNMVKNLTSQMYEYGOVRGE 291
Qy 188 LGVGLQDLSGLQNS--YDNKEGAVVISVEKDSAPKAGILVWDLITVEVNGKVKNTNEL 245
Db 292 LGIMGTLSNELAKAMKYDAQRGAFVSQVMPNSSAAKAGIKAGDVITSLNGKPISSFAAL 351
Qy 246 RNLIGSMPLPNORVTLKVIIRDKKERAFTLTLAERKNPNKKTISAQNGAQQGLNGQVDEL 305
Db 352 RAQVGTMPVGSKISLGLREGK--AITVNL-----ELQQSSQSQVDSSTIFSG 397
Qy 306 TQETKRSRLSDVGVVLVSQVNSPAPQAGFRGNIITKIEEVEYKSVADFNALEKY 365
Db 398 IEGAEKSNKGD--KGVVVSSVKANSAPAAQIGLKKGDVIGANQIPVKNIAEIRKILD-- 453
Qy 366 KGKPKRFLVLDLNOG 380

Db 454 -SKPS-VIALNIORG 466

Search completed: November 18, 2002, 11:30:31
Job time : 19 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:33:46 ; Search time 18 Seconds
(without alignments)
632.593 Million cell updates/sec

Title: US-09-895-913A-120
Perfect score: 387
Sequence: 1 MIPKERMALSGVILSKD.....KPKRFLVLDLNGGYRIILVK 387

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	3.1	463	1 US-08-485-569-2	Sequence 2, Appli
2	12	3.1	463	1 US-08-480-993-2	Sequence 2, Appli
3	12	3.1	463	2 US-07-903-079B-2	Sequence 2, Appli
4	12	3.1	472	1 US-08-278-091-5	Sequence 5, Appli
5	12	3.1	472	1 US-08-483-859-5	Sequence 5, Appli
6	12	3.1	472	1 US-08-472-173-5	Sequence 5, Appli
7	12	3.1	472	2 US-08-487-167-5	Sequence 5, Appli
8	12	3.1	472	2 US-08-482-816-5	Sequence 5, Appli
9	12	3.1	472	2 US-08-296-149-5	Sequence 5, Appli
10	12	3.1	472	2 US-08-801-499-5	Sequence 5, Appli
11	12	3.1	472	3 US-08-615-271-5	Sequence 5, Appli
12	12	3.1	472	3 US-09-074-660-5	Sequence 5, Appli
13	12	3.1	472	3 US-09-074-659-5	Sequence 5, Appli
14	12	3.1	472	3 US-09-106-468-5	Sequence 5, Appli
15	12	3.1	472	4 US-09-106-466A-5	Sequence 5, Appli
16	12	3.1	472	4 US-09-106-467-5	Sequence 5, Appli
17	12	3.1	475	1 US-08-278-091-6	Sequence 6, Appli
18	12	3.1	475	1 US-08-483-859-6	Sequence 6, Appli
19	12	3.1	475	1 US-08-472-173-6	Sequence 6, Appli
20	12	3.1	475	1 US-08-350-741-2	Sequence 2, Appli
21	12	3.1	475	2 US-08-487-167-6	Sequence 6, Appli
22	12	3.1	475	2 US-08-482-816-6	Sequence 6, Appli
23	12	3.1	475	2 US-08-296-149-6	Sequence 6, Appli
24	12	3.1	475	2 US-08-801-499-6	Sequence 6, Appli
25	12	3.1	475	2 US-08-463-875A-2	Sequence 2, Appli
26	12	3.1	475	2 US-08-615-271-6	Sequence 6, Appli
27	12	3.1	475	3 US-09-074-660-6	Sequence 6, Appli

28	12	3.1	475	3 US-09-074-659-6	Sequence 6, Appli
29	12	3.1	475	3 US-09-106-468-6	Sequence 6, Appli
30	12	3.1	475	4 US-09-106-466A-6	Sequence 6, Appli
31	12	3.1	475	4 US-09-106-467-6	Sequence 6, Appli
32	12	3.1	491	3 US-08-923-454A-14	Sequence 14, Appli
33	12	3.1	492	1 US-08-350-741-3	Sequence 3, Appli
34	12	3.1	492	2 US-08-463-875A-3	Sequence 3, Appli
35	11	2.8	503	1 US-08-245-294-8	Sequence 8, Appli
36	11	2.8	503	1 US-08-474-499-8	Sequence 8, Appli
37	11	2.8	503	1 US-08-307-279A-8	Sequence 8, Appli
38	11	2.8	503	4 US-09-525-310-8	Sequence 8, Appli
39	11	2.6	503	5 PCT-US95-06211-8	Sequence 35, Appli
40	10	2.6	238	4 US-09-605-858-35	Sequence 35, Appli
41	10	2.6	413	4 US-09-134-001C-5111	Sequence 5111, Ap
42	10	2.6	580	4 US-08-818-112-75	Sequence 75, Appli
43	10	2.6	580	4 US-08-818-111-76	Sequence 76, Appli
44	10	2.6	580	4 US-09-058-556-75	Sequence 75, Appli
45	10	2.6	580	4 US-09-072-596-76	Sequence 76, Appli

ALIGNMENTS

RESULT 1
US-08-485-569-2
; Sequence 2, Application US/08485569
; Patent No. 5679547
; GENERAL INFORMATION:
; APPLICANT: Krivan, Howard C.
; APPLICANT: Samuel, James E.
; TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
; TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,569
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/903,079
; FILING DATE: 22-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,966
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/631,698
; FILING DATE: 21-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

US-08-485-569-2

Query Match 3.1%; Score 12; DB 1; Length 463;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 TNNHVIDGADKI 36

|||||

Db 114 TNNHVIDGADKI 125

RESULT 2

US-08-480-993-2

; Sequence 2, Application US/08480993

; Patent No. 5721115

; GENERAL INFORMATION:

; APPLICANT: Krivan, Howard C.

; APPLICANT: Samuel, James E.

; TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE

; TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/480,993

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/903,079

; FILING DATE: 22-JUN-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/810,966

; FILING DATE: 20-DEC-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/631,698

; FILING DATE: 21-DEC-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7969-050

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 463 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; US-08-480-993-2

Query Match 3.1%; Score 12; DB 1; Length 463;

Best Local Similarity 100.0%; Pred. No. 0.00067;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 TNNHVIDGADKI 36

|||||

Db 114 TNNHVIDGADKI 125

RESULT 3

US-07-903-079B-2

; Sequence 2, Application US/07903079B

; Patent No. 5843463

; GENERAL INFORMATION:

; APPLICANT: Krivan, Howard C.

; APPLICANT: Samuel, James E.

; TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE

; TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/903,079B

; FILING DATE: 22-JUN-1992

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/810,966

; FILING DATE: 20-DEC-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/631,698

; FILING DATE: 21-DEC-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7969-004

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 463 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; US-07-903-079B-2

Query Match 3.1%; Score 12; DB 2; Length 463;

Best Local Similarity 100.0%; Pred. No. 0.00067;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 TNNHVIDGADKI 36

|||||

Db 114 TNNHVIDGADKI 125

RESULT 4

US-08-278-091-5

; Sequence 5, Application US/08278091

; Patent No. 5506139

; GENERAL INFORMATION:

; APPLICANT: LOOSMORE, Sheena M

; APPLICANT: YANG, Yan-Ping

; APPLICANT: CHONG, Pele

; APPLICANT: OOMEN, Raymond P.

; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with

; TITLE OF INVENTION: Reduced Protease Activity

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/278,091
FILING DATE: 21-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-371
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-278-091-5

Query Match 3.1%; Score 12; DB 1; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

160 GIGFAIPSNMVK 171
|||||
Db 263 GIGFAIPSNMVK 274

RESULT 5
US-08-483-859-5
Sequence 5, Application US/08483859
Patent No. 5656436
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: COHEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
REDUCED PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,859
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091

FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-859-5
Query Match 3.1%; Score 12; DB 1; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 160 GIGFAIPSNMVK 171
|||||
Db 263 GIGFAIPSNMVK 274
RESULT 6
US-08-472-173-5
Sequence 5, Application US/08472173
Patent No. 5665353
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: COHEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
REDUCED PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,173
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-472-173-5

Query Match 3.1%; Score 12; DB 1; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
|||||
Db 263 GIGFAIPSNMVK 274

RESULT 7

US-08-487-167-5

; Sequence 5, Application US/08487167
; Patent No. 5869302

; GENERAL INFORMATION:

; APPLICANT: LOOSMORE, Sheena M.

; APPLICANT: YANG, Yan-Ping

; APPLICANT: CHONG, Pele

; APPLICANT: OOMEN, Raymond P.

; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with

; TITLE OF INVENTION: Reduced Protease Activity

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/487,167

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/296,149

; FILING DATE: 26-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/278,091

; FILING DATE: 21-JUL-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 472 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-487-167-5

Query Match

3.1%; Score 12; DB 2; Length 472;

Best Local Similarity 100.0%; Pred. No. 0.00068;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
|||||
Db 263 GIGFAIPSNMVK 274

RESULT 8

US-08-482-816-5

; Sequence 5, Application US/08482816

;

;

;

;

;

;

COUNTRY: Canada
ZIP: MSG IN7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
ATTORNEY/AGENT INFORMATION:
CLASSIFICATION: 514
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-390
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-296-149-5

Query Match 3.1%; Score 12; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
|||||
DB 263 GIGFAIPSNMVK 274

RESULT 10
US-08-801-499-5
; Sequence 5, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG IN7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,499
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-671 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-801-499-5

Query Match 3.1%; Score 12; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
|||||
DB 263 GIGFAIPSNMVK 274

RESULT 11
US-08-615-271-5
; Sequence 5, Application US/08615271
; Patent No. 5981503
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; TITLE OF INVENTION: PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG IN7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-615-271-5

Query Match 3.1%; Score 12; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
|||||

Db 263 GIGFAIPSNMVK 274

RESULT 12

US-09-074-660-5
; Sequence 5, Application US/09074660
; Patent No. 6020183
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-731 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-074-660-5

Query Match 3.1%; Score 12; DB 3; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
Db 263 GIGFAIPSNMVK 274

RESULT 13

US-09-074-659-5
; Sequence 5, Application US/09074659
; Patent No. 6025342
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele

11-11-11-11-11

; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,659
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-730 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-074-659-5

Query Match 3.1%; Score 12; DB 3; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
Db 263 GIGFAIPSNMVK 274

RESULT 14

US-09-106-468-5
; Sequence 5, Application US/09106468
; Patent No. 6114125
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada

ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106.468
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/615,271
FILING DATE: 20-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-825
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-106-468-5

Query Match 3.1%; Score 12; DB 3; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 160 GIGFAIPSNMVK 171
Db 263 GIGFAIPSNMVK 274

RESULT 15
US-09-106-466A-5
Sequence 5, Application US/09106466A
Patent No. 6147057
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
TITLE OF INVENTION: PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106.466A
FILING DATE:
CLASSIFICATION: 514514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/615,271
FILING DATE: 20-JUN-1996
CLASSIFICATION: 514514
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-826
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-106-466A-5
Query Match 3.1%; Score 12; DB 4; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 160 GIGFAIPSNMVK 171
Db 263 GIGFAIPSNMVK 274

Search completed: November 18, 2002, 11:36:20
Job time : 19 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:19:35 ; Search time 39 Seconds
(without alignments)
1322.256 Million cell updates/sec

Title: US-09-895-913a-120
Perfect score: 1938
Sequence: 1 MIPKMERALSGVVISKD.....KPKRFLVLDLNOGYRIILVK 387

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	1938	100.0	387 19 AAW98445	H. pylori GHPO 536
2	1934	99.8	443 20 AAW89850	Antigen 1 from clu
3	1904	98.2	451 20 AAW89977	Protein encoded by
4	1441	74.4	288 20 AAW89984	Expressed antigen
5	909	46.9	187 20 AAW89817	Protein encoded by
6	749.5	38.7	491 20 AAV37309	Protein involved i
7	722.5	37.3	488 23 ABB90581	Chlamydia pneumoni
8	718.5	37.1	491 19 AAW56771	Escherichia coli s
9	714.5	36.9	475 12 AAR14625	S. typhimurium Htr
10	684	35.3	474 22 AAG78605	Lawsonia intracell

11	683	35.2	433	21	AAV59353	BASB011 protein se
12	683	35.2	433	21	AAV59354	BASB011 protein se
13	683	35.2	433	21	AAV59355	BASB011 protein se
14	681.5	35.2	460	20	AAV29294	Protein encoded by
15	666	34.4	463	17	AAV2690	H. influenzae SB33
16	663	34.2	437	21	AAAB1834	Haemophilus influe
17	658	34.0	437	21	AAAB1837	Haemophilus influe
18	658	34.0	437	21	AAAB1838	Haemophilus influe
19	657	33.9	437	21	AAAB1836	Haemophilus influe
20	656	33.8	437	21	AAAB1835	Haemophilus influe
21	652	33.6	455	22	AAV98946	E. coli growth and
22	616.5	31.8	503	17	AAV77434	Heat shock protein
23	603	31.1	499	21	AAV52993	Neisseria meningit
24	601	31.0	499	21	AAV52994	Neisseria meningit
25	599	30.9	499	21	AAV75749	Neisseria meningit
26	599	30.9	499	21	AAV52995	Neisseria meningit
27	594	30.7	448	21	AAV70409	Neisseria meningit
28	594	30.7	499	21	AAV75750	Neisseria meningit
29	590	30.4	499	21	AAV75748	Neisseria gonorrhoe
30	584.5	30.2	475	21	AAV70413	Neisseria meningit
31	584.5	30.2	498	21	AAV70413	Neisseria meningit
32	559.5	28.9	414	21	AAV83151	NGSP polypeptide o
33	559.5	28.9	465	21	AAV83150	NGSP polypeptide o
34	536.5	27.7	413	23	ABP40266	Staphylococcus epi
35	534.5	27.6	317	22	AAV82365	S. epidermidis ope
36	528	27.2	370	21	AAV52996	Neisseria meningit
37	520.5	26.9	355	22	AAV98947	E. coli growth and
38	501	25.9	344	22	AAV60173	Propionibacterium
39	480.5	24.8	500	23	ABV48818	Listeria monocytog
40	476.5	24.6	317	20	AAV35634	Chlamydia pneumoni
41	475	24.5	409	21	AAV91334	Group B Streptococ
42	474.5	24.5	448	23	ABV93659	Herbicide daily activ
43	474	24.5	380	23	ABP27963	Streptococcus poly
44	474	24.5	409	23	ABP29849	Streptococcus poly
45	474	24.5	409	23	ABP30135	Streptococcus poly

ALIGNMENTS

RESULT 1

AAW98445
ID AAW98445 standard; Protein; 387 AA.

XX AAW98445;

AC AAW98445;

DT 31-MAR-1999 (first entry)

XX H. pylori GHPO 536 protein.

DE GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

KW peptic ulcer disease.

XX Helicobacter pylori.

OS Helicobacter pylori.

XX WO9843478-A1.

PD 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

PF 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX14164...

PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
XX infections and gastrointestinal diseases
XX
PS Claim 8; Page 769-771; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 387 AA;

Query Match 100.0%; Score 1938; DB 19; Length 387;
Best Local Similarity 100.0%; Pred. No. 3.8e-149;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MIPKMERALGSGVVIISKDGYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSDDL 60
DB 1 MIPKMERALGSGVVIISKDGYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSDDL 60
QY 61 AVIRITKDNLPITKIFSDSDNDISVGLVFAIGNPFGVGSVTGIVSALNKGIGINSYEN 120
DB 61 AVIRITKDNLPITKIFSDSDNDISVGLVFAIGNPFGVGSVTGIVSALNKGIGINSYEN 120
QY 121 FIQTASINPGSGGALIDSRGLVGINTAIISKGTGNGHIGIFAIPSNMVKDVTQLIKT 180
DB 121 FIQTASINPGSGGALIDSRGLVGINTAIISKGTGNGHIGIFAIPSNMVKDVTQLIKT 180
QY 181 GKIERGYLGVLQDLSGDLQNSYDNKEGAVISVEKDSAPKAGILVWDLITEVNGKKVK 240
DB 181 GKIERGYLGVLQDLSGDLQNSYDNKEGAVISVEKDSAPKAGILVWDLITEVNGKKVK 240
QY 241 NTNELNLIGSMLPNQRTVLKVIROKKEAFTLTAEKRNPNKKTETISQAQAGQOLNGL 300
DB 241 NTNELNLIGSMLPNQRTVLKVIROKKEAFTLTAEKRNPNKKTETISQAQAGQOLNGL 300
QY 301 QVEDLTQETKRSMLSDDDVQGVLSQVNSNPAEQAGFRQGNITIKIEVEVKSADFNNH 360
DB 301 QVEDLTQETKRSMLSDDDVQGVLSQVNSNPAEQAGFRQGNITIKIEVEVKSADFNNH 360
QY 361 ALEKYKGPKEFLVLDLNOGYRIILVK 387
DB 361 ALEKYKGPKEFLVLDLNOGYRIILVK 387

RESULT 2
AAW89850
TD AAW89850 standard; Protein; 443 AA.
XX
XX AAW89850;
XX
XX
XX 18-FEB-1999 (first entry)
XX
XX Antigen 1 from cluster 04a.
XX
XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
OS Helicobacter pylori.
XX
XX WO9849314-A2.
XX
XX 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08487.
XX
XX 14-OCT-1997; 97US-0061958.
XX
XX 25-APR-1997; 97US-0045107.
XX

PA (GENE-) GENELABS TECHNOLOGIES INC.
XX Chow TP, Fry KE, Lim MY, McAtee CP;
XX WPI; 1999-009433/01.
XX
XX New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
XX
PS Claim 1; Page 211-212; 402pp; English.
XX
CC The present sequence represents a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ Sequence 443 AA;

Query Match 99.8%; Score 1934; DB 20; Length 443;
Best Local Similarity 99.7%; Pred. No. 9.7e-149;
Matches 386; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MIPKMERALGSGVVIISKDGYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSDDL 60
DB 57 MIPKMERALGSGVVIISKDGYIVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSDDL 116
QY 61 AVIRITKDNLPITKIFSDSDNDISVGLVFAIGNPFGVGSVTGIVSALNKGIGINSYEN 120
DB 117 AVIRITKDNLPITKIFSDSDNDISVGLVFAIGNPFGVGSVTGIVSALNKGIGINSYEN 176
QY 121 FIQTASINPGSGGALIDSRGLVGINTAIISKGTGNGHIGIFAIPSNMVKDVTQLIKT 180
DB 177 FIQTASINPGSGGALIDSRGLVGINTAIISKGTGNGHIGIFAIPSNMVKDVTQLIKT 236
QY 181 GKIERGYLGVLQDLSGDLQNSYDNKEGAVISVEKDSAPKAGILVWDLITEVNGKKVK 240
DB 237 GKIERGYLGVLQDLSGDLQNSYDNKEGAVISVEKDSAPKAGILVWDLITEVNGKKVK 296
QY 241 NTNELNLIGSMLPNQRTVLKVIROKKEAFTLTAEKRNPNKKTETISQAQAGQOLNGL 300
DB 297 NTNELNLIGSMLPNQRTVLKVIROKKEAFTLTAEKRNPNKKTETISQAQAGQOLNGL 356
QY 301 QVEDLTQETKRSMLSDDDVQGVLSQVNSNPAEQAGFRQGNITIKIEVEVKSADFNNH 360
DB 357 QVEDLTQETKRSMLSDDDVQGVLSQVNSNPAEQAGFRQGNITIKIEVEVKSADFNNH 416
QY 361 ALEKYKGPKEFLVLDLNOGYRIILVK 387
DB 417 ALEKYKGPKEFLVLDLNOGYRIILVK 443

RESULT 3
AAW89977
ID AAW89977 standard; Protein; 451 AA.
XX
XX AAW89977;
XX
XX 18-FEB-1999 (first entry)
XX
XX Protein encoded by clone b8.
XX
XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
OS Helicobacter pylori.

```
XX  W09849314-A2.
XX
XX
XX  05-NOV-1998.
XX
XX  27-APR-1998; 98WO-US08487.
XX
XX  14-OCT-1997; 97US-0061958.
XX  25-APR-1997; 97US-0045107.
XX
XX  (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX  Chow TP, Fry KE, Lim MY, McAtee CP;
XX  WPI; 1999-009433/01.
XX  N-PSDB; AAV90921.
XX
XX  New Helicobacter pylori antigens and related nucleic acid sequences
XX  - useful in serological diagnosis and protective vaccines, providing
XX  long-lasting immune response
XX
XX  Claim 15; Page 317-318; 402pp; English.
XX
XX  The present sequence represents a Helicobacter pylori antigenic protein
XX  that is characterised by immunoreactivity with H. pylori-positive
XX  antisera. The proteins are highly immunogenic and induce a long-lasting
XX  immune response that persists even after antimicrobial treatment. In
XX  antibody-detection assays, on sera, plasma, urine, saliva etc., they are
XX  highly sensitive and specific. The specification also describes 69
XX  previously unrecognised immunogenic cluster families. H. pylori antigens
XX  are used to detect H. pylori-specific antibodies, for diagnosing
XX  infection or to confirm eradication of infection, and in vaccines to
XX  protect against H. pylori infection and related diseases (gastritis,
XX  peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX  Sequence 451 AA;
XX
XX  Query Match 98.2%; Score 1904; DB 20; Length 451;
XX  Best Local Similarity 98.4%; Pred. No. 2.7e-146;
XX  Matches 381; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
XX
Qy 1 MIPKRMERALGSGVIISKDGYIYVNNHVIDGADKIKVTIPGSNKYSATLVGTDSDDL 60
Db 65 MIPKRMERALGSGVIISKDGYIYVNNHVIDGADKIKVTIPGSNKYSATLVGTDSDDL 124
Qy 61 AVIRTKDNLTIFKSDNDISVGLVFAIGNPRGVGSVTQGIIVSALNKGSGINSYEN 120
Db 125 AVIRTKDNLTIFKSDNDISVGLVFAIGNPRGVGSVTQGIIVSALNKGSGINSYEN 184
Qy 121 FIQTDAINPNSGALIDSRGGLVGINTAIISKTGGNHGIGFAIPSNMVKDVTVTLTKT 180
Db 185 FIQTDAINPNSGALIDSRGGLVGINTAIISKTGGNHGIGFAIPSNMVKDVTVTLTKT 244
Qy 181 GKIERGYLGVLQDLSDGLQNSYDNKEGAVVISEKDSAPAKAGILVWDLTIVNGKKVK 240
Db 245 GKIERGYLGVLQDLSDGLQNSYDNKEGAVVISEKDSAPAKAGILVWDLTIVNGKKVK 304
Qy 241 NTNELNLIGSMLPNQRTLVIRDKKERAFVTLTIAERKNPKKETISAQNGAOGQNLG 300
Db 305 NTNELNLIGSMLPNQRTLVIRDKKERAFVTLTIAERKNPKKETISAQNGAOGQNLG 364
Qy 301 QVEDLTQTKRSMRLSDVQGVLYSQVNENSPAEQAGFRQGNITKIEVEVKSVDADPNH 360
Db 365 QVEDLTQTKRSMRLSDVQGVLYSQVNENSPAEQAGFRQGNITKIEVEVKSVDADPNH 424
Qy 361 ALEKYKGPKRFLVDLNLQGYRIILVK 387
Db 425 ALEKYKGPKRFLVDLNLQGYRIILVK 451
XX
XX  RESULT 4
XX  AAW89984
XX  ID AAW89984 standard; Protein: 288 AA.
```

```
XX  AAW89984;
XX
XX  18-FEB-1999 (first entry)
XX
XX  Expressed antigen for clone Y261A.
XX
XX  Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX  peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
XX  Helicobacter pylori.
XX
XX  W09849314-A2.
XX
XX  05-NOV-1998.
XX
XX  27-APR-1998; 98WO-US08487.
XX
XX  14-OCT-1997; 97US-0061958.
XX  25-APR-1997; 97US-0045107.
XX
XX  (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX  Chow TP, Fry KE, Lim MY, McAtee CP;
XX  WPI; 1999-009433/01.
XX
XX  New Helicobacter pylori antigens and related nucleic acid sequences
XX  - useful in serological diagnosis and protective vaccines, providing
XX  long-lasting immune response
XX
XX  Claim 16; Page 322-323; 402pp; English.
XX
XX  The present sequence represents a Helicobacter pylori antigenic protein
XX  that is characterised by immunoreactivity with H. pylori-positive
XX  antisera. The proteins are highly immunogenic and induce a long-lasting
XX  immune response that persists even after antimicrobial treatment. In
XX  antibody-detection assays, on sera, plasma, urine, saliva etc., they are
XX  highly sensitive and specific. The specification also describes 69
XX  previously unrecognised immunogenic cluster families. H. pylori antigens
XX  are used to detect H. pylori-specific antibodies, for diagnosing
XX  infection or to confirm eradication of infection, and in vaccines to
XX  protect against H. pylori infection and related diseases (gastritis,
XX  peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX  Sequence 288 AA;
XX
XX  Query Match 74.4%; Score 1441; DB 20; Length 288;
XX  Best Local Similarity 100.0%; Pred. No. 6.1e-109;
XX  Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 100 VTQGIIVSALNKGSGIGINSYENFIQTDASINPNSGALIDSRGGLVGINTAIISKTGGNH 159
Db 1 VTQGIIVSALNKGSGIGINSYENFIQTDASINPNSGALIDSRGGLVGINTAIISKTGGNH 60
Qy 160 GIGFAIPSNMVKITVTQLIKTKIERGYLGVLQDLSDGLQNSYDNKEGAVVISEKDS 219
Db 61 GIGFAIPSNMVKITVTQLIKTKIERGYLGVLQDLSDGLQNSYDNKEGAVVISEKDS 120
Qy 220 AKKAGILVWDLTIVNGKKVKNTNELNLIGSMLPNQRTLVIRDKKERAFVTLTIAERK 279
Db 121 AKKAGILVWDLTIVNGKKVKNTNELNLIGSMLPNQRTLVIRDKKERAFVTLTIAERK 180
Qy 280 NPNKKTISAQNGAOGQNLQGVLYSQVNENSPAEQAGFRQGNITKIEVEVKSVDADPNH 339
Db 181 NPNKKTISAQNGAOGQNLQGVLYSQVNENSPAEQAGFRQGNITKIEVEVKSVDADPNH 240
Qy 340 QGNITKIEVEVKSVDADPNHNALEKYKGPKRFLVDLNLQGYRIILVK 387
Db 241 QGNITKIEVEVKSVDADPNHNALEKYKGPKRFLVDLNLQGYRIILVK 288
XX
XX  RESULT 5
```

AAW89817
ID AAW89817 standard; Protein; 187 AA.
XX AC AAW89817;
XX DT 18-FEB-1999 (first entry)
XX DE Protein encoded by clone C7.
XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX OS Helicobacter pylori.
XX PN W09849314-A2.
XX PD 05-NOV-1998.
XX PF 27-APR-1998; 98WO-US08487.
XX PR 14-OCT-1997; 97US-0061958.
XX PR 25-APR-1997; 97US-0045107.
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX Chow TP, Fry KE, Lim MY, McAtee CP;
XX WPI; 1999-009433/01.
XX DR N-PSDB; AAW90552.
XX PT New Helicobacter pylori antigens and related nucleic acid sequences
XX PT - useful in serological diagnosis and protective vaccines, providing
XX PT long-lasting immune response
XX PS Claim 15; Page 103; 402pp; English.
XX CC The present sequence represents a Helicobacter pylori antigenic protein
XX CC that is characterised by immunoreactivity with H. pylori-positive
XX CC antisera. The proteins are highly immunogenic and induce a long-lasting
XX CC immune response that persists even after antimicrobial treatment. In
XX CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
XX CC highly sensitive and specific. The specification also describes 69
XX CC previously unrecognised immunogenic cluster families. H. pylori antigens
XX CC are used to detect H. pylori-specific antibodies, for diagnosing
XX CC infection or to confirm eradication of infection, and in vaccines to
XX CC protect against H. pylori infection and related diseases (gastritis,
XX CC peptic ulcer, gastric adenocarcinoma/lymphoma).
XX SQ Sequence 187 AA;
Query Match 46.9%; Score 909; DB 20; Length 187;
Best Local Similarity 98.4%; Pred. No. 5.6e-66;
Matches 182; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 203 YDNKEGAVVISVEKDSPPAKKAGILVWDLITEVNGKKVNTNLRNLISMLPNQRTLVK 262
DB 3 YDNKEGAVVISVEKDSPPAKKAGILVWDLITEVNGKKVNTNLRNLISMLPNQRTLVK 62
QY 263 IRDKKRAFTLTAERKPNKKTISQAQGLNGQLQVEDLTQTKRSMRLSDVQGV 322
DB 63 IRDKKRAFTLTAERKPNKKTISQAQGLNGQLQVEDLTQTKRSMRLSDVQGV 122
QY 323 LVSQVNSPBAQFROGNIITKIEEVEKSVADFNHALEKYKPKRFLVLDLNOGYR 382
DB 123 LVSQVNSPBAQFROGNIITQIEEVEKSVADFNHALEKYKPKRFLVLDLNOGYR 182
QY 383 IILVK 387
DB 183 IILVK 187
RESULT 6
AAW37309

AAW37309 standard; Protein; 491 AA.
XX AC AAW37309;
XX DT 07-OCT-1999 (first entry)
XX DE Protein involved in intermediate metabolism of polypeptides.
XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
XX KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX OS Chlamydia trachomatis.
XX PN W09928475-A2.
XX PD 10-JUN-1999.
XX PF 27-NOV-1998; 98WO-IB01939.
XX PR 04-NOV-1998; 98US-0107077.
XX PR 28-NOV-1997; 97ER-0015041.
XX PR 17-DEC-1997; 97ER-0016034.
XX (GEST) GENSET.
XX Griffais R;
XX WPI; 1999-371125/31.
XX PT Genome sequence of Chlamydia trachomatis
XX PS Disclosure; Page 1044-1045; 1755pp; English.
XX CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
XX CC of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
XX CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
XX CC can also be used to control growth of the microorganism. Chlamydia
XX CC trachomatis is responsible for a large number of diseases, e.g. eye
XX CC diseases such as conventional trachoma, nonendemic trachoma,
XX CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
XX CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
XX CC perihhepatitis, bartholinitis; pneumopathy in breast feeding infants;
XX CC and venereal lymphogranulomatosis. The polypeptides of the invention
XX CC may be of use in treating these diseases.
XX SQ Sequence 491 AA;
Query Match 38.7%; Score 749.5; DB 20; Length 491;
Best Local Similarity 41.5%; Pred. No. 2.1e-52;
Matches 162; Conservative 83; Mismatches 126; Indels 19; Gaps 7;
QY 3 PKRERALGSGVILSKDGYVTNNHVIDGADKIKVTIPGSNKKEYSATLVGCTDSGLAY 62
DB 113 PQOR-DAVRGTGFVSDGYVTNNHVIDGADKIKVTIPGSNKKEYSATLVGCTDSGLAY 170
QY 63 IRITKDNLTPTKFSNDISVDGLVFAIGNPFGVSVTQGVISALNKGSGIGINSYENFI 122
DB 171 IKIQAEKLPFLTFGNSDQLOIGDWAIGTGNPGLQATVTVGVVSAKGRNQLHIVDFEDI 230
QY 123 QTDASINPNSGGLIDSRGLVGINATISKTGNGHIGFAIPSNMVKDVTYTLIKTK 182
DB 231 QTDAAINPNSGGLINQGVIGVNTAIVSGSGYIGIGFAIPSLMAKRVLDLISDQ 290
QY 183 IERGVGLVGLDLSGLQNSY--DNKEGAVISVEKDSPPAKKAGILVWDLITEVNGKKV 240
DB 291 VTRGFLGVTLPQIDSELATCYKLEKYGALVTDVVKGSPAERQEDVIVAYNGKEVE 350
QY 241 NTNELNLIGSLPNQRTLVKIRKRAFTLTLAERKNPNKKTISQAQGLNGQLNG 299
DB 351 SUSALRNAISLMPGTRVILKIVREGKTIETPTVTQ-----IPTEDGVSALQKNG 401

Qy 300 LOVEDLTQETKSRMLSDDDVQGLVSVQVNSPABQAGFRQGNITTKIEVEVKSVADFN 359
Db 402 VVVQNTPEICKKGLAATGILVAVAGSPASAGVAPCQLILAVNRQVRASVEELN 461
Qy 360 HALEYKGPKEKRFVLVDLNOG--YRILVK 387
Db 462 QVLKNSRGEN---VLLMVSGQDVVRFIVIK 488

RESULT 7
ABB90581
ID ABB90581 standard; Protein; 488 AA.
XX AC ABB90581;
XX
DT 29-JUL-2002 (first entry)
DE Chlamydia pneumoniae cp7306 protein, SEQ ID NO:111.
XX
XX Chlamydia pneumoniae cp7306 protein, SEQ ID NO:111.
KW Chlamydia pneumoniae cp7306 protein, SEQ ID NO:111.
KW human respiratory disease; cardiovascular disease; atherosclerosis;
KW coronary artery disease; carotid artery stenosis; myocardial infarction;
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW strain CWL029.
XX
OS Chlamydia pneumoniae.
XX
XX
FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein /label= Signal_peptide
FT 21..488
FT /note= "Mature protein"
XX
XX W0200202606-A2.
XX
PD 10-JAN-2002.
XX
XX 03-JUL-2001; 2001WO-IB01445.
XX
XX 03-JUL-2000; 2000GB-0016363.
XX 11-JUL-2000; 2000GB-0017047.
XX 21-JUL-2000; 2000GB-0017983.
XX 07-AUG-2000; 2000GB-0019368.
XX 18-AUG-2000; 2000GB-0020440.
XX 14-SEP-2000; 2000GB-0022583.
XX 10-NOV-2000; 2000GB-0027549.
XX 22-DEC-2000; 2000GB-0031706.
XX
PA (CHIR-) CHIRON SPA.
XX
XX Ratti G, Grandi G;
XX
XX WPI; 2002-154726/20.
XX N-PSDB; ABL91239.
XX
XX Novel Chlamydia pneumoniae protein useful in the manufacture of a
XX medicament for treatment or prevention of infection due to Chlamydia,
XX preferably Chlamydia pneumoniae, and for diagnostic purposes
XX
XX Claim 1; Page 95; 364pp; English.
XX
XX Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
XX pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
XX them. The proteins are predicted to be immunogenic and may therefore be
XX useful in vaccine production and for diagnostic purposes. Chlamydia
XX pneumoniae is a common cause of respiratory disease in humans, and is
XX also involved in the development of cardiovascular diseases such as
XX atherosclerosis, coronary artery disease, carotid artery stenosis,
XX myocardial infarction, cerebrovascular disease, aortic aneurysm,
XX claudication and stroke. The proteins and nucleic acids of the invention
XX may be used in vaccines and pharmaceutical compositions for the
XX prevention or treatment of chlamydial infections, particularly Chlamydia
XX pneumoniae infections. The proteins may also be used in the detection of

CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC DNA probe assay or blotting techniques for determining Chlamydia
CC pneumoniae gene expression. The present sequence represents a
CC specifically claimed Chlamydia pneumoniae protein of the invention.
XX
SQ Sequence 488 AA;
Query Match 37.3%; Score 722.5; DB 23; Length 488;
Best Local Similarity 40.0%; Pred. No. 3.3e-50;
Matches 158; Conservative 82; Mismatches 134; Indels 21; Gaps 7;
Qy 2 IPKER-----MERALGSGVIISKDGIYIVTNHVIDGADKIKVTPGPSNKEYSATLVGTD 56
Db 103 LPSOREKPSQSKAVRGTCFLVSPDGIYIVTNHVVEDTKIHRVTLHDGOK-YPATVIGLDP 161
Qy 57 ESDLAVIRITKDNLPITKFSNDSISVGLVFAIGNPFGVSGVSTQGIIVSALNKGSGIGN 116
Db 162 KTDLAIVIKIKSONLPYLSFGNSDHLKVGDMATAIGNPFLQATVTVGVISAKRNQLHIA 221
Qy 117 SYENFIQTDASINPGNSGGALIDSRGGLVGINTAIISKTGNGHIGFAIPSMNVKDTVTQ 176
Db 222 DFEDFIQDAAINPGNSGGPLNIDGQVIGVNTAIVSGGGYIGIGFAIPSLMANRIIDQ 281
Qy 177 LIKTKIERGYLGVGLDGLQNSY--DNKEGAVTISVEKDSAPKAGILVWDLITEV 234
Db 282 LTRDQVTRGFLVTLQPIDAELAAACYKLEKIVGALVTDVVKSPADKAGLKQEDVIAY 341
Qy 235 NGKKVKNTNELNLIGSMLPNORVTLKVIKDKKRAFTLTAEKRNPNKKTETISAQNGAQ 294
Db 342 NGKVDLSUSMFRNAVSLMNPDPTRIVLKVREGKVIETVTVSQAB---KEDGMSAL---- 394
Qy 295 GQLNGLOVEDLTQETKSRMLSDDDVQGLVSVQVNSPABQAGFRQGNITTKIEVEVK 354
Db 395 -QRVGIRVQNTPEICKKGLAATGILVAVAGSPASAGVAPCQLILAVNRQVRASVE 453
Qy 355 VADFNHALEKYKGPKEKRFVLVDLNOG--YRILVK 387
Db 454 IEDLNRTL---KDSNNENILMVSGQDVIRFIALK 485

RESULT 8
AAW56771
ID AAW56771 standard; Protein; 491 AA.
XX AC AAW56771;
XX
DT 13-OCT-1998 (first entry)
XX
XX Escherichia coli serine protease htrA.
XX
XX PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease; htrA;
XX serine protease; neurodegeneration; predisposition; diagnosis.
XX
XX Escherichia coli.
XX
XX EP828003-A2.
XX
XX 11-MAR-1998.
XX
XX 26-AUG-1997; 97HP-0306501.
XX
XX 13-DEC-1996; 96US-0032875.
XX 06-SEP-1996; 96US-0025436.
XX 25-OCT-1996; 96US-0027873.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
XX
XX Browne MJ, Clinkenbeard HE, Creasy CL, Karran EH;
XX Livi GP, Southan CD;
XX WPI; 1998-161101/15.
XX N-PSDB; AAV29528.
XX

```
XX Nucleic acids encoding human serum protease protein(s) - used for
PT diagnosing pre-disposition to Alzheimer's disease, etc.
XX
PS Example 2; Page 28-29; 65pp; English.
XX
CC The sequence is that of the serine protease htrA which was used
CC in the isolation of pSP1.
XX
SQ Sequence 491 AA;
Query Match 37.1%; Score 718.5; DB 19; Length 491;
Best Local Similarity 43.2%; Pred. No. 7e-50;
Matches 163; Conservative 73; Mismatches 116; Indels 25; Gaps 9;
QY 10 ALGSGVIIKQD-GYIVTNHHVIDGADKIKVTIPGSGNKEYSATLVGTDSESLAVIRI-TK 67
DB 112 ALGSGVIIDADKGYVVTNNHHVDNATVIVQL-SGGRKFDKMWKDPKSDIALIQINP 170
QY 68 DNLPTIKFSDSDISVGDILVFAIGNPFGYGESVTOGIVSALNKSIGIGINSYENFIQTDA 127
DB 171 KNLTAIKMADSDALRVGDYTVGIGNPFGIGETVTSIGIVSALGRSGLNAENYENFIQTDA 230
QY 128 INPNSGGALIDSRGLVGINTAISKTCGNHGIGFAIPSNMVKDTVTOLIKTGKIERGY 187
DB 231 INRNSGGALVNLGELIGINTAILAPDGGNIGIGFAIPSNMVKNTLSQMEYGVYKRG 290
QY 188 LGVGLQDLGSDLQNS--VDNKEGAVVISVEKDSPAKAGILVWDLITEVNGKVKNTNEL 245
DB 291 LGIMTELSELAKAMKVDQAGFAVSQVLPNSRAKAGIKAGDVITSUNGKPISSFAAL 350
QY 246 RNLIGSMLPNQRTLVKVRIDKKERAFTLTLAERKNPNKKTETISQAQAGQQLNGLQVEDL 305
DB 351 RAQVGTMPVGSKLTLGLLRDQGVNVLQPQ--SSQNVQDSSSIFNGIBG----- 399
QY 306 TOETKRSMLSDVQGVLSQVNSPAPQAGFROGNIITKITEEVKSVADFNHALEY 365
DB 400 ---AEMSNKGD--QGVVYNNYKTGTPTAAQIGLKKGDVIIGANQQAVKNIABLKVLD--- 452
QY 366 KGKPKRFLVLDLNOGYR 382
DB 453 -SKPS-VLALNIQGRD 467
RESULT 9
AARI4625
ID AARI4625 standard; Protein; 475 AA.
XX
AC AARI4625;
XX
DT 14-JAN-1992 (first entry)
XX
S. typhimurium HtrA protein.
XX
KW TnpHoA mutagenesis; signal peptide; stress protein;
KW heat shock protein; degP.
XX
OS Salmonella typhimurium strain C5.
XX
PN WO9115572-A.
XX
PD 17-OCT-1991.
XX
PF 23-MAR-1991; 91WO-GB00484.
XX
PR 30-MAR-1990; 90GB-0007194.
XX
PA (WELL ) WELLCOME FOUNDATION LTD.
XX
PI Dougan G, Charles IG, Hormaeche CE, Johnson KS, Chatfield SN;
XX WPI; 1991-325215/44.
XX DR N-PSDB; AAI414416.
```

```
XX Attenuated microorganism useful in live vaccines - attenuated by
PT mutation in DNA sequence encoding e.g. a heat shock protein
XX
PS Disclosure; Fig 1; 28pp; English.
XX
CC TnpHoA mutagenesis was used in the mouse virulent S. typhimurium
CC strain C5. Mutants were selected likely to harbour lesions in genes
CC that have a signal peptide sequence, i.e. proteins likely to be
CC targeted through a bacterial membrane. Isolation of the DNA
CC flanking the TnpHoA insertion identified the gene (htrA) that has
CC been insertionally activated. Comparison of the translated protein
CC sequence showed that it shared 88% homology with the sequence of the
CC htrA product from E. coli.
XX
SQ Sequence 475 AA;
Query Match 36.9%; Score 714.5; DB 12; Length 475;
Best Local Similarity 42.9%; Pred. No. 1.4e-49;
Matches 161; Conservative 75; Mismatches 114; Indels 25; Gaps 9;
QY 10 ALGSGVII-SKDGIVTNNHHVIDGADKIKVTIPGSGNKEYSATLVGTDSESLAVIRI-TK 67
DB 113 ALGSGVIIDAAKGYVVTNNHHVDNATVIVQL-SGGRKFDKMWKDPKSDIALIQINP 171
QY 68 DNLPTIKFSDSDISVGDILVFAIGNPFGYGESVTOGIVSALNKSIGIGINSYENFIQTDA 127
DB 172 KNLTAIKLADSDALRVGDYTVGIGNPFGIGETVTSIGIVSALGRSGLNENYENFIQTDA 231
QY 128 INPNSGGALIDSRGLVGINTAISKTCGNHGIGFAIPSNMVKDTVTOLIKTGKIERGY 187
DB 232 INRNSGGALVNLGELIGINTAILAPDGGNIGIGFAIPSNMVKNTLSQMEYGVYKRG 291
QY 188 LGVGLQDLGSDLQNS--VDNKEGAVVISVEKDSPAKAGILVWDLITEVNGKVKNTNEL 245
DB 292 LGIMTELSELAKAMKVDQAGFAVSQVLPNSRAKAGIKAGDVITSUNGKPISSFAAL 351
QY 246 RNLIGSMLPNQRTLVKVRIDKKERAFTLTLAERKNPNKKTETISQAQAGQQLNGLQVEDL 305
DB 352 RAQVGTMPVGSKLTLGLLRDQGVNVLQPQ--SSQNVQDSSSIFNGIBG----- 397
QY 306 TOETKRSMLSDVQGVLSQVNSPAPQAGFROGNIITKITEEVKSVADFNHALEY 365
DB 398 IEQAEMSNKGD--KGVVYSSVKANSPPAAQIGLKKGDVIIGANQQAVKNIABLKVLD--- 453
QY 366 KGKPKRFLVLDLNOG 380
DB 454 -SKPS-VLALNIQRG 466
RESULT 10
AAG78605
ID AAG78605 standard; Protein; 474 AA.
XX
AC AAG78605;
XX
DT 20-NOV-2001 (first entry)
XX
DE Lawsonia intracellularis protein SEQ ID NO: 7.
XX
KW HtrA; PonA; HypC; Yefw; ABC1; Omp100; Lawsonia intracellularis infection;
KW vaccine.
XX
OS Lawsonia intracellularis.
XX
PN JP2001169787-A.
XX
PD 26-JUN-2001.
XX
PF 20-OCT-2000; 2000JP-0320736.
XX
PR 22-OCT-1999; 99US-0160922.
XX
```


[illegible]

RESULT 11	
AA59353	
ID	AA59353 standard; Protein; 433 AA.
XX	
AC	AA59353;
XX	
DT	10-MAR-2000 (first entry)
XX	
DE	BASB011 protein sequence #1.
XX	
BW	BASB011; immune response; bacterial infection; middle ear infection;
KW	bacterial adhesion; extracellular matrix protein; tissue damage; therapy.
XX	
OS	Moraxella catarrhalis.
XX	
PN	WO955871-A1.
XX	
PD	04-NOV-1999.
XX	
PF	20-APR-1999; 99WO-EP02764.
XX	
PR	23-APR-1998; 98GB-0008720.
XX	

```

(SWIK ) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;
WPI: 2000-062033/05.
N-PSDB: AAZ48700.

New polypeptides from Moraxella catarrhalis used to treat the infection
by this bacteria -
Claim 3; Fig 3; 168pp; English.

This sequence represents a BAS011 protein from Moraxella catarrhalis.
The polypeptide is used to generate an immune response in an animal,
particularly against a bacterial infection, e.g. a Moraxella catarrhalis
infection. M. catarrhalis is present in 15% of childhood middle ear
infections in the US. Molecules of the invention may also be used to
prevent adhesion of bacteria to extracellular matrix proteins on
indwelling devices or in wounds, to block bacterial adhesion between
extracellular matrix proteins and BAS011 proteins that mediate tissue
damage, or to block the normal progression of pathogenesis in infections
initiated other than by implanting of indwelling devices or by other
surgical techniques.

Sequence 433 AA;

Query Match 35.2%; Score 683; DB 21; Length 433;
Best Local Similarity 45.3%; Pred. No. 4.5e-47;
Matches 136; Conservative 73; Mismatches 83; Indels 8; Gaps

Qy 11 LGSQVIISKDGYIVTNHHVIDGADKIKVTIPGSKNYSATLVGTDSSESLAVIRITKDL 70
Db 131 LGSQVVVSSEGYVTNAHVINGADEITVAL-NDGRRKARATVIGSDADSLAVIKVELDNL 189

Qy 71 PTIKFSNDNISVGDLVFAIGNPFGVGSYSTQGIIVSALNKGSGIGINSYENFTQDASINP 130
Db 190 VPMFAFR-AEPIRVGDVSLAIGCNPFVGQTVTQGIISATGTGTGIGVSFEDFIQTDAAINP 248

Qy 131 GNSGGLIDSRGLVGLNTALISKTGNGHGTGPAIPSNMVKDVTOLIKTKTERGYIGV 190
Db 249 GNSGGLVDANGALIGINTALYSKSGSGMGIGPAIPNQIVQQVMTSLITGTGKVSROWMGI 308

Qy 191 GLQDLSGDLQNSYDNKSGEVAIVSEKDSKPAKKAGILVWDLITVEVNGKVKNTNELRNLC 250
Db 309 EMVRMTDDPTN-LESRNSVIRRWQNSPAEHLKSGDKIVRIDGVHITSINELVGVA 367

Qy 251 SMLPNQRVTLKVLRDKKERAFITLTLAEKPNPNKKTETIS--AQNGAQQQLNGQVEDLTQE 308
Db 368 RKAPDSOLITVEINRDORPMTVOVILARRPS---SETLSQPVNSPSQSRSTOLEQLOE 424

```

RESULT 12	
AAV59354	
ID	AAV59354 standard; Protein; 433 AA.
XX	
AC	AAV59354;
XX	
DT	10-MAR-2000 (first entry)
XX	
DE	BASB011 protein sequence #2.
XX	
KW	BASB011; immune response; bacterial infection; middle ear infection;
KW	bacterial adhesion; extracellular matrix protein; tissue damage; therapy.
XX	
OS	Moraxella catarrhalis.
XX	
PN	WO9955871-A1.
XX	
PD	04-NOV-1999.
XX	
PF	20-APR-1999; 99WO-EP02764.
XX	
PR	23-APR-1998; 98CB-0008720.
XX	

```
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Ruelle J;
XX PI WPI: 2000-062033/05.
XX DR N-PSDB; AA248701.
XX PT New polypeptides from Moraxella catarrhalis used to treat the infection
XX PT by this bacteria -
XX PS Claim 3; Fig 3; 108pp; English.
XX CC This sequence represents a BAS011 protein from Moraxella catarrhalis.
XX CC The polypeptide is used to generate an immune response in an animal,
XX CC particularly against a bacterial infection, e.g. a Moraxella catarrhalis
XX CC infection. M. catarrhalis is present in 15% of childhood middle ear
XX CC infections in the US. Molecules of the invention may also be used to
XX CC prevent adhesion of bacteria to extracellular matrix proteins on
XX CC indwelling devices or in wounds, to block bacterial adhesion between
XX CC extracellular matrix proteins and BAS011 proteins that mediate tissue
XX CC damage, or to block the normal progression of pathogenesis in infections
XX CC initiated other than by implanting of indwelling devices or by other
XX CC surgical techniques.
XX SQ Sequence 433 AA;
Query Match 35.2%; Score 683; DB 21; Length 433;
Best Local Similarity 45.3%; Pred. No. 4.5e-47;
Matches 136; Conservative 73; Mismatches 83; Indels 8; Gaps 5;
QY 11 LGSGLVTSKDGIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSESLAVIRITKDNL 70
DB 131 LGSGLVTSKDGIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSESLAVIRITKDNL 189
QY 71 PTIKFSDSDNDISVGLDLVFAIGNPFGVGSVTQGIIVSALNKSIGINSYENFIQTDAINP 130
DB 190 VPMAFR-AEPIRVGDVSLAIGNPFGVGSVTQGIIVSALNKSIGINSYENFIQTDAINP 248
QY 131 GNSGGALIDSRGLVGINTAISKTGNGHIGFAIPSNMVKDVTOLIKTKIERGYLGV 190
DB 249 GNSGGALVDPANGALIGINTAIYSRSGSGMGIGFAIPNQIVQVMTSLITITGKVSQWMI 308
QY 191 GLQDLSGLDQNSVDNKGAVISVEKDSAPKAGILVWDLITEVNGKKVNTNELRLIG 250
DB 309 EMVRMTDDPTN-IESRSNVIIRRVQNSPAEHAGLKSQKIVRIDGVHITSINELVGVA 367
QY 251 SMLPNQVTLKVRDKKERAFTLTIAERKNPNKKTIS--AQNGAQCQLNGLOVEDLTQE 308
DB 368 RKAPDSQLTVEIMRDQRPMTVQVILAERPS---SETLSQPVQNSPSQSRQSTQLEQLLQE 424
RESULT 13
ID AAY59355 standard; Protein; 433 AA.
XX AC AAY59355;
XX DT 10-MAR-2000 (first entry)
XX DE BAS011 protein sequence #3.
XX KW BAS011; immune response; bacterial infection; middle ear infection;
XX KW bacterial adhesion; extracellular matrix protein; tissue damage; therapy.
XX OS Moraxella catarrhalis.
XX PN WO9955871-A1.
XX PD 04-NOV-1999.
XX PF 20-APR-1999; 99WO-EP02764.
XX PF 20-APR-1999; 99WO-EP02764.
PR 23-APR-1998; 98GB-0008720.
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PI Ruelle J;
PI WPI: 2000-062033/05.
DR N-PSDB; AA248702.
DR PT New polypeptides from Moraxella catarrhalis used to treat the infection
DR PT by this bacteria -
DR PS Claim 3; Fig 3; 108pp; English.
DR CC This sequence represents a BAS011 protein from Moraxella catarrhalis.
DR CC The polypeptide is used to generate an immune response in an animal,
DR CC particularly against a bacterial infection, e.g. a Moraxella catarrhalis
DR CC infection. M. catarrhalis is present in 15% of childhood middle ear
DR CC infections in the US. Molecules of the invention may also be used to
DR CC prevent adhesion of bacteria to extracellular matrix proteins on
DR CC indwelling devices or in wounds, to block bacterial adhesion between
DR CC extracellular matrix proteins and BAS011 proteins that mediate tissue
DR CC damage, or to block the normal progression of pathogenesis in infections
DR CC initiated other than by implanting of indwelling devices or by other
DR CC surgical techniques.
DR SQ Sequence 433 AA;
Query Match 35.2%; Score 683; DB 21; Length 433;
Best Local Similarity 45.3%; Pred. No. 4.5e-47;
Matches 136; Conservative 73; Mismatches 83; Indels 8; Gaps 5;
QY 11 LGSGLVTSKDGIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSESLAVIRITKDNL 70
DB 131 LGSGLVTSKDGIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSESLAVIRITKDNL 189
QY 71 PTIKFSDSDNDISVGLDLVFAIGNPFGVGSVTQGIIVSALNKSIGINSYENFIQTDAINP 130
DB 190 VPMAFR-AEPIRVGDVSLAIGNPFGVGSVTQGIIVSALNKSIGINSYENFIQTDAINP 248
QY 131 GNSGGALIDSRGLVGINTAISKTGNGHIGFAIPSNMVKDVTOLIKTKIERGYLGV 190
DB 249 GNSGGALVDPANGALIGINTAIYSRSGSGMGIGFAIPNQIVQVMTSLITITGKVSQWMI 308
QY 191 GLQDLSGLDQNSVDNKGAVISVEKDSAPKAGILVWDLITEVNGKKVNTNELRLIG 250
DB 309 EMVRMTDDPTN-IESRSNVIIRRVQNSPAEHAGLKSQKIVRIDGVHITSINELVGVA 367
QY 251 SMLPNQVTLKVRDKKERAFTLTIAERKNPNKKTIS--AQNGAQCQLNGLOVEDLTQE 308
DB 368 RKAPDSQLTVEIMRDQRPMTVQVILAERPS---SETLSQPVQNSPSQSRQSTQLEQLLQE 424
RESULT 14
ID AAY29294 standard; Protein; 460 AA.
XX AC AAY29294;
XX DT 25-OCT-1999 (first entry)
XX DE Protein encoded by the PA14 degP gene.
XX KW Human pathogen; virulence polypeptide; virulence factor;
XX KW pathogenic infection; Pseudomonas aeruginosa infection.
XX OS Pseudomonas aeruginosa.
XX PN WO9927129-A1.
XX PD 03-JUN-1999.
XX PF 25-NOV-1998; 98WO-US25247.
```

```
XX 25-NOV-1997; 97US-0066517.
XX (GEO ) GEN HOSPITAL CORP.
XX Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
XX Rahne LG, Tan M, Tsongalis J;
XX WPI; 1999-357851/30.
XX Virulence factors useful in developing disease treatments
XX Disclosure; Fig 28; 228pp; English.
XX The present sequence represents a Pseudomonas aeruginosa polypeptide
XX sequence. P. aeruginosa is an opportunistic human pathogen present in
XX soil water and plants. The specification describes virulence polypeptides
XX and nucleic acid sequence encoding such polypeptides. These sequences
XX can be used to identify a compound which is capable of decreasing the
XX expression of a pathogenic virulence factor. Compounds that inhibit
XX the expression or activity of virulence factor polypeptides can be
XX used to treat pathogenic infections, especially where the infection
XX is a P. aeruginosa infection.
XX note: the sequences given in the specification were poorly legible, and
XX in some instances assumptions were made as to the identity of the
XX residue; it is therefore possible that the sequence given below is
XX not entirely correct.
XX Sequence 460 AA;
XX Query Match 35.2%; Score 681.5; DB 20; Length 460;
XX Best Local Similarity 41.3%; Pred. No. 6.5e-47;
XX Matches 157; Conservative 72; Mismatches 132; Indels 19; Gaps 8;
QY 4 KERMERALGSGVILSKDGYIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTDSEDLAVI 63
DB 90 QOREAQSLGGGFIISMDGYIILTNHVVADAEILVRL-SDRSEHKAKLVGADPRSDVAVL 148
QY 64 RITKDNLTPIKFSNDISVGLDVFALGNPFGVGSVTVQGISVSNLKNKSGIGINSYENFIQ 123
DB 149 KEAKNLPTLKLGDMSKLYGKVLAIQSPFGDHSVTAGIVSAKGRS-LPMESVVPFIQ 207
QY 124 TDASTINPNSGGALIDSRGVLGINTAIISKTGGNHGTGFAIPSNMVKDVTVQLIKTKGI 183
DB 208 TDVAIPNPGSGPLNLNLEGEVVGINSQIFTRSGGFMGLSFAIPIDVALNVADQLKKGKV 267
QY 184 ERGYLGVGLQDLGDLQNSY--DNKEGAVVISVEKDSPAKKAGILVWDLITEVNGKVKVN 241
DB 268 SHGVLGVVQIEVKNKDLAESFGLDKPSGALVAQLVEDGPAAGGLQVGVVLSLNGQSINE 327
QY 242 TNELNLIGSMLPNQRTVLKVRDKKERAFITLTAERKNPNKKTISAQNGAQQGLN--- 298
DB 328 SADLPHLVGMKPGDKINLDVIRGQRK--SLSMVAGNLPDDDEI-ASMGAPAESRM 384
QY 299 --GLOVEDLTQETKSRMLSDDDVQGVVLVSQVNSPESQAGFRQGNITIKIEVEVKSIV- 355
DB 385 RLGVTVADLTAEQRKSL---DIQGVVVIKEVQDGPAAVIGLRPGDVITHLDMRAVTSK 440
QY 356 --ADFNHALEYKPKRFL 373
DB 441 IFADVAKALPKMRSVSMRVL 460
RESULT 15
AAR92690
ID AAR92690 standard; Protein; 463 AA.
XX AC AAR92690;
XX 03-JUN-1996 (first entry)
XX H. influenzae SB33 Hin47.
XX
```

```
KW Hin47; adhesin; protease; mutagenesis; vaccine; diagnosis;
KW meningitis; epiglottitis, pneumonia; otitis media.
XX Haemophilus influenzae strain SB33.
XX WO9603506-A2.
XX 08-FEB-1996.
XX 21-JUL-1995; 95WO-CA00434.
XX 07-JUN-1995; 95US-0487167.
XX 21-JUL-1994; 94US-0278091.
XX 26-AUG-1994; 94US-0296149.
XX (CONN-) CONNAUGHT LAB LTD.
XX Chong P, Klein MH. Loosmore SM, Oomen RP, Yang Y;
XX WPI; 1996-117051/12.
XX N-PSDB; AAT10423.
XX H.influenzae Hin47 protein analogue with decreased protease activity
XX - and nucleic acid encoding it, useful in vaccines and as diagnostic
XX agents
XX Example 2; Fig 2A-2H; 75pp; English.
XX Outer membrane protein Hin47 (AAR92690) of H. influenzae SB33 is
XX reported to be an adhesin, and is conserved among strains of H.
XX influenzae. For Hin47 to be used in vaccines, its protease
XX activity must be decreased to prevent degradation of other antigens.
XX This was performed by site-directed mutagenesis of the hin47 gene
XX (AAT10423). Hin47 analogues in which e.g. Ser197 was replaced by Ala,
XX Cys or Thr, His31 by Ala, Lys or Arg, or Asp121 by Ala were produced
XX by transformed host cells. The analogues had less than 10% of the
XX protease activity of native Hin47. They can also be used to raise
XX antibodies and as carriers in prodn. of conjugate vaccines.
XX Sequence 463 AA;
XX Query Match 34.4%; Score 666; DB 17; Length 463;
XX Best Local Similarity 41.0%; Pred. No. 1.2e-45;
XX Matches 150; Conservative 75; Mismatches 117; Indels 24; Gaps 8;
QY 5 ERMERALGSGVIT-SKCGYIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTDSEDLAVI 63
DB 93 KRNFRGLGSGVITINASKGYIVLTNNHVIDEADKITVQLQ-DGREFAKLVLGKDESLDALV 151
QY 64 RITK-DNLPTIKFSNDISVGLDVFALGNPFGVGSVTVQGISVSNLKNK-SGIGINSYENF 121
DB 152 QLEKPSNLTEIKFADSKLVRGDFVAILGNPFGLGQTVTSIGIVSALGRSTGSDSGYENY 211
QY 122 IOTDASINPNSGGALIDSRGVLGINTAIISKTGGNHGTGFAIPSNMVKDVTVQLIKTG 181
DB 212 IOTDAAVNRGNSGGALVNLGELIGINTAIISPSGNGAGIAFAIPSNQASNLVQOILEFG 271
QY 182 KIERGYLGVGLQDLGDLQNSYD--NKEGAVVISVEKDSPAKKAGILVWDLITEVNGKV 239
DB 272 QVRRGLLGKGGFLNADLAKAFNVSAQQGAFVSEVLPKSAEAKGLKAGDIITAMNGOKI 331
QY 240 KNTNELNLIGSMLPNQRTVLKVRDKKERAFITLTA--ERKNPNKKTISAQNGAQQGL 297
DB 332 SSFAEIRAKIATGAGKKEISLTLRLDGKSHDKVKKMLQADSSQLSSKTELP------L 384
QY 298 NGLQVEDLTQETKSRMLSDDDVQGVVLVSQVNSPESQAGFRQGNITIKIEVEVKSIVAD 357
DB 385 DGATLKQYD-----AKGVKGIETIKIQPNSLAAQRLKSGDIIIGINRQMIENIRE 435
QY 358 FNHALE 363
DB 436 LNKVLE 441
```

Search completed: November 18, 2002, 11:28:36
Job time : 41 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:20:03 ; Search time 14 Seconds
(without alignments)
1146.524 Million cell updates/sec

Title: US-09-895-913a-120

Perfect score: 1938

Sequence: 1 MIPKMERALGSGVIISKD.....KPKRFLVLDNQYRIILVK 387

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	749.5	38.7	497	1	DEGP_CHLTR
2	746.5	38.5	497	1	DEGP_CHLMO
3	722.5	37.3	474	1	DEGP_ECOLI
4	722.5	37.3	488	1	DEGP_CHLPN
5	714.5	36.9	475	1	DEGP_SALTY
6	679.5	35.1	513	1	DEGP_BRUB
7	679	35.0	466	1	HTOA_BACIN
8	678.5	35.0	478	1	DEGP_BUCAI
9	663.5	34.2	508	1	DEGP_RICCN
10	657.5	33.9	478	1	DEGP_BUCAP
11	652	33.6	455	1	DEGP_ECOLI
12	625.5	32.3	513	1	DEGP_RICPR
13	618.5	31.9	504	1	DEGP_RHIME
14	616.5	31.8	503	1	DEGP_BARHE
15	554.5	28.6	340	1	DEGS_HAEIN
16	520.5	26.9	355	1	DEGS_ECOLI
17	508	26.2	458	1	YVTA_BACSU
18	493	25.4	449	1	HTRA_BACSU
19	477	24.6	437	1	DEGL_ARATH
20	474.5	24.5	448	1	DEGS_ARATH
21	451.5	23.3	400	1	YVXA_BACSU
22	439	22.7	413	1	HTRA_LACHE
23	414.5	21.4	408	1	HTRA_LACLA
24	380.5	19.6	480	1	HRAL_HUMAN
25	380.5	19.6	480	1	HRAL_MOUSE
26	357.5	18.4	453	1	HRA3_HUMAN
27	348.5	18.0	458	1	HRA2_MOUSE
28	340.5	17.6	476	1	HRA2_HUMAN
29	333.5	17.2	458	1	HRA2_HUMAN
30	305	15.7	321	1	HHOA_ARATH
31	304	15.7	460	1	HRA3_MOUSE
32	226.5	11.7	630	1	Y4BJ_RHISN
33	143	7.4	280	1	ETA_STAAU

RESULT 1
DEGP_CHLTR STANDARD; PRT; 497 AA.
AC PI8584; O84830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa immunogenic protein) (SK59).
DE DEGP OR HTRA OR Ctr823.
GN Chlamydia trachomatis.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar L2;
RX MEDLINE=90337348; PubMed=2379836;
RA Kahane S., Weinstein I., Sarov I.;
RT "Cloning, characterization and sequence of a novel 59-kDa protein of Chlamydia trachomatis.";
RL Gene 90:61-67(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=99008009; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -|- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -|- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001355; AAC68420.1; -;
CC EMBL; M31119; AAA23116.1; -;
CC PHCI-2DPAGE; P18584; -;
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001940; Protease2C.
CC InterPro; IPR00126; Ser_proteas_v8.
CC InterPro; IPR001254; Ser_protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC Pfam; PF00595; PDZ; 2.
CC PRINTS; PR00834; PROTEASES2C.

P53930 saccharomyc
O35435 mus musculu
P25937 salmonella
P22252 campylobact
Q53634 staphylococ
P47715 mycoplasma
P22251 campylobact
Q01681 p genome po
P33666 escherichia
P17767 p genome po
P13529 p genome po

ALIGNMENTS

```
DR PRINTS; PRO0839; V8PROTEASE.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Repeat; Signal; Antigen;
KW Complete proteome.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 497 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 128 289 CATALYTIC.
FT DOMAIN 290 381 PDZ 1.
FT DOMAIN 394 485 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 497 AA; 53244 MW; 86A5E31BB8A438BA CRC64;

Query Match
Best Local Similarity 41.3%; Score 749.5; DB 1; Length 497;
Matches 161; Conservative 84; Mismatches 126; Indels 19; Gaps 7;

OY 3 PKRMERALGSGVIISKDGIYVNNHVIDGADKIKVTIPGSNKEYSATLVGTDSEDLAV 62
Db 119 PQOR-DAVRGTGFIYSEDGYVVTNNHHVVEDAGKIHVTLHDGQK-YTAKIVGLDPKTDLAV 176
OY 63 IRTKDNLTIFKSDNISVDGLVFAIGNPFGVSGVTOGIVSALNKSIGINSYENFI 122
Db 177 IKIQAEKLPFLTFGNSDQLQIGDWAIGNPFGLOATVTVGVISAKGRNLHIVDFEDI 236
OY 123 QTDASINPGNSGGLIDSRGGLVGINTAISKTGGNHGIGFAIPSNMVKDVTYQLIKTKG 182
Db 237 QTDAAINPGNSGGLIDSRGGLVGINTAISKTGGNHGIGFAIPSNMVKDVTYQLIKTKG 296
OY 183 IERGVLGVLQDLSGLQNSY--DNKEGAVVISVEKDSPPAKKAGILVMDLITEVNGKVK 240
Db 297 VTRGFLGVTLPIDSELAACYKLEKRYGALITDVVKGSPAERAGLRQEDVIYVAYNGKEVE 356
OY 241 NTNELNLIGSMLPNQRTVLKIRDKKERAFTLTIAERKNPNKKTETISAQAGQ-QLNG 299
Db 357 SUSALRNAISLMPGTRVVKLVKVRREGKTEIPVTYQ-----IPTEDGVSALQKMG 407
OY 300 LOVEDLTQETKSRMLSDDDQVGLVSOVNSPAGQAFROGNIITKIEVEVKSVADFN 359
Db 408 VRVQNTPEICKKGLAADTRGILVAVANGSPAASAGVAPQGLILAVNRQVRASVEELN 467
OY 360 HALEKYKPKRFLVLDLNOG--YRIILVK 387
Db 468 QVLKNSKGEN---VLLMWSGQDVVRIVLK 494

RESULT 2
DEGP_CHLMU STANDARD; PRT; 497 AA.
AC Q9PL97;
OY 16-OCT-2001 (Rel. 40, Created)
OY 16-OCT-2001 (Rel. 40, Last sequence update)
OY 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR TC0210.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR EGS0165.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htrA gene of Escherichia
RT coli: a sigma 32-independent mechanism of heat-inducible
RT transcription.";
RL Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Okanishi M., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90323597; PubMed=2165018;
RA Quirk S., Bhatnagar S.K., Bessman M.J.;
RT "Primary structure of the deoxyguanosine triphosphate
RT triphosphohydrolase-encoding gene (dgt) of Escherichia coli.";
RL Gene 89:13-18(1990).

RN [8]
RP SEQUENCE OF 1-16 FROM N.A.
RX MEDLINE=90207273; PubMed=2157212;
RA Wurgler S.M., Richardson C.C.;
RT "Structure and regulation of the gene for dGTP triphosphohydrolase
RT from Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
RN [9]
RP CHARACTERIZATION, AND SEQUENCE OF 27-39.
RX MEDLINE=90202693; PubMed=2180903;
RA Lipinska B., Zyllicz M., Georgopoulos C.;
RT "The HtrA (DegP) protein, essential for Escherichia coli survival at
RT high temperatures, is an endopeptidase.";
RL J. Bacteriol. 172:1791-1797(1990).
RN [10]
RP IDENTITY OF HTRA AND PROTEASE DO.
RX MEDLINE=91222240; PubMed=2025286;
RA Seol J.H., Woo S.-K., Jung E.M., Yoo S.-J., Lee C.S., Kim K.J.,
RA Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
RT "Protease Do is essential for survival of Escherichia coli at high
RT temperatures: its identity with the htrA gene product.";
RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE
CC ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGP.
CC -!- SUBUNIT: MULTIMERIC.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- INDUCTION: BY HEAT SHOCK.
CC -!- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X36536; AAA33994.1; -
CC EMBL; X36537; CAA30997.1; -
CC EMBL; D26562; BRA05608.1; -
CC EMBL; AE000125; AAC73272.1; -
CC EMBL; U70214; AAB08591.1; -
CC EMBL; AE005192; AAG54465.1; -
CC EMBL; AP002550; BAB33588.1; -
CC EMBL; M29955; AAA33717.1; -
CC EMBL; M31772; AAA33680.1; -
CC PIR; S01899; S01899.
CC PIR; B35993; B35993.
CC MEROPS; S01.273; -
CC SWISS-2DPAGE; P09376; COLI.
CC EcoGene; EG10463; degP.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001940; Protease2C.
CC InterPro; IPR001254; Ser.protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC Pfam; PF00595; PDZ; 2.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS50106; PDZ; 2.
KW Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 474 PROTEASE DO.
FT DOMAIN 280 371 PDZ 1.
FT DOMAIN 377 466 PDZ 2.
FT ACT_SITE 131 131 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 161 161 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 236 236 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 10 10 A -> R (IN REF. 1, 7 AND 8).

```
FT CONFLICT 46 46 E -> Q (IN REF. 7).
FT CONFLICT 192 192 A -> G (IN REF. 1).
FT CONFLICT 467 474 STYILIMQ -> RHLPVNAVISLNPFLKTGRGSPYNL (IN
FT REF. 1).
SQ SEQUENCE 474 AA: 49354 MW: 5482E596F74B6D5F CRC64;

Query Match 37.3%; Score 722.5; DB 1; Length 474;
Best Local Similarity 43.5%; Pred. No. 1.7e-35;
Matches 163; Conservative 74; Mismatches 113; Indels 25; Gaps 9;

QY 10 ALGSGVLIISKD-CYIVTNHNVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRI-TK 67
DB 112 ALGSGVLIIDADGQYVVTNNHVDNATVIKVL-SDGRKFDAKXMGKDPDSIALIOIQNP 170
QY 68 DNPPTIKFSDSDISVGLDVFALGNFPGVSGVTCQIVSALNKGSGIGINSYENFIOTDAS 127
DB 171 KNLTAIKMADSDALRVGDYTVALGNFPGVSGVTCQIVSALNKGSGIGINSYENFIOTDAA 230
QY 128 INPGSGGALIDSRGLVGINTAISKGTGNHGHGIGFAIPSNMVKDPVTOLIKTKIERGY 187
DB 231 INRNGSGGALVNLGELIGINTAILAPDGGNIGIGFAIPSNMVKNTLSQWVEYGVQVRGE 290
QY 188 LGVGLDLSDDLQNS--YDNKEGAVVISVEKDSPAKAGILVWDLITEVNGKKVKTNEL 245
DB 291 LGIMGTLSLSELAKAMKVDQAGAFVQVLPNPSAAKAGIKAGDVITSLNGKPISSFAAL 350
QY 246 RNLIGSMPLNQRTLVKIVRDKKERAFTLTLAERKNPNKKTETISAQNGAQLNGLQVEDL 305
DB 351 RAQVGTPMVGSKLTGLLRDQGVNVNLEL-QOSSNQVDSISFNGIEG----- 399
QY 306 TQETKSRMLSDVQGVLVSVQNVNSPABQAGFRQGNITKIEEVEKSVADFNHALEYK 365
DB 400 ---AEMSNKRGKD--QGVVNVNKTGTPAAQIGLKGKGVIIIGANQQAKVNAELRKVLD-- 452
QY 366 KGKPKRFLVLDLNOG 380
DB 453 -SKPS-VLALNIQRG 465

RESULT 4
ID DEGP_CHLPN STANDARD; PRT: 488 AA.
AC Q926T0; Q9JQD7; Q9K1W4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR CPN0979 OR CP0877.
OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaeae; Chlamydia phila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CWL029;
RA MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RA "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RA Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-AR39;
RA MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RA pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001678; AAD19116.1; -.
DR EMBL; AE002246; AAF38665.1; -.
DR EMBL; AP002548; BAA99186.1; -.
DR MEROPS; S01.273; -.
DR PHCI-2DPAGE; Q9Z6T0; -.
DR TIGR; CP0877; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 488 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 119 280 CATALYTIC.
FT DOMAIN 281 372 PDZ 1.
FT DOMAIN 388 476 PDZ 2.
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 164 164 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 488 AA: 52311 MW: 0EE7E0F8F106F49 CRC64;

Query Match 37.3%; Score 722.5; DB 1; Length 488;
Best Local Similarity 40.0%; Pred. No. 1.7e-35;
Matches 158; Conservative 82; Mismatches 134; Indels 21; Gaps 7;

QY 2 IPKER-----MERALGSGVLIISKDGYIVTNHNVIDGADKIKVTIPGSKNKEYSATLVGTD 56
DB 103 LPSQREKPOSKEAVRGTFGLVSPDGYIVTNHNVEDTGKIHVTHDQK-YPATVIGLDP 161
QY 57 ESDLAVIRITKDNLPITKFSNDSNDISVGLDVFALGNFPGVSGVTCQIVSALNKGSGIGIN 116
DB 162 KYDLAVIKIKSNLPYLSFGNSDHLKVGWATAIGNPFGLQATVTVGVISAKGNQLHTA 221
QY 117 SYENFIOTDASINPGSGGALIDSRGLVGINTAISKGTGNHGHGIGFAIPSNMVKDVTQ 176
DB 222 DEEDFIQTDAAINPGSGGALLNIDQGVIVNTAIVSGSGYIGIGFAIPSLMANRIQD 281
QY 177 LIKTGKIERGYLGVLQDLSDDLQNSY--DNKEGAVVISVEKDSPAKAGILVWDLITEV 234
DB 282 LIRDGVTRGFLGVTLOPDAELACYKLEKVGALVTDVWKGSPADKAGLQEDYIIAY 341
QY 235 NGCKYKNTNELNLGSLMPLNORVTLKVRDKKERAFTLTLAERKNPNKKTETISAQNGA 294
DB 342 NKEVDLSLNFNRNLSLMPDTRIVLVVREGKVIPIPTVSQAP---KEDGMSAL----- 394
QY 295 GOLNGLQVEDLTAQEKRSRLSDVQGVLVSVQNVNSPABQAGFRQGNITKIEEVEKSV 354
DB 395 -QRVGIRVQNLTPETAKKGIAPETKGIILISVEPSSVAASSGAPGLILAVNRQKVS 453
QY 355 VADFNHAELEYKGPKRFLVLDLNOG--YRILVK 387
```


Db 454 IEDLNTL---KDSNNENILLMVSGDVIRFTALK 485

DEGP_SALTY

RESULT 5

ID DEGP_SALTY STANDARD; PRT; 475 AA.

AC P26982;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DE Protease do precursor (EC 3.4.21.-).

GN DEGP OR HTRA OR PTD OR STM0209.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Salmonella.

CC NCBI_TaxID=602;

CC [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C5;

RX MEDLINE=91251770; PubMed=1645840;

RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,

RA Ali T., Miller I., Hormaeche C.;

RT "The role of a stress-response protein in Salmonella typhimurium

RT virulence.";

RL Mol. Microbiol. 5:401-407(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=LT2 / SGSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

RT LT2.";

RL Nature 413:852-856(2001).

CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.

CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED

CC SPECIFICITY WITH HHOA/DEGO.

CC -!- SUBCELLULAR LOCATION: Periplasmic.

CC -!- INDUCTION: BY HEAT SHOCK.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.

CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X54548; CAA38420.1; -.

CC EMBL; AE008704; AAL19173.1; -.

CC PIR; S15337; S15337.

CC MIP; S21327; S21327.

CC MEROPS; S01.273; -.

CC StyGene; SG10173; degp.

CC InterPro; IPR001478; degp.

CC InterPro; IPR001940; Protease2C.

CC InterPro; IPR001254; Ser_protease_Try.

CC Pfam; PF00089; trypsin; 1.

CC Pfam; PF00595; PDZ; 2.

CC PRINTS; PR00834; PROTEASES2C.

CC SMART; SM00228; PDZ; 2.

CC PROSITE; PS0106; PDZ; 2.

CC Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;

CC Complete proteome.

CC SIGNAL 1 26

CC CHAIN 27 475

CC DOMAIN 281 372

CC DOMAIN 378 467

CC

CC BY SIMILARITY.

CC PROTEASE DO.

CC PDZ 1.

CC PDZ 2.

CC

FT FT

FT FT


```
Qy 358 FNHALE 363
Db 439 LNKVLE 444

RESULT 8
DEGP_BUCAI STANDARD; PRT; 478 AA.
AC P57322;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR BU228.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
SS symbiotic bacterium).
SC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RL Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
Qy 182 KIERGYLGVLQQLSGDLQN--SYDNKKGAVVISVEKDSPAKKAGILYWDLITEYNGKV 239
Db 287 QVRGELGIMGELNSDLAQIMKINSQKGAFSRVLPNNSAFEAGIKAGDIIISLNRKPI 346
Qy 240 KNTNELRNIGSMLPNQVRVTLKIRDKKERAPFTLAEKKNPNKKETISAOAGQQLNG 299
Db 347 SFSFSLRAEIGSLPVATKRMELGVFREGRTKNTITVELKH----SVKHNLSNENDYIG-IEG 401
Qy 300 LOVEDLTQETKREMLSDDDVGVLVSQVNSPAGQAGFROGNIITKTTEEVEVKSVADEN 359
Db 402 VLSDY-----IFNEQKVIKVDNVKPHTPASKIGFKDDIILNINQKLISNVE-- 450
Qy 360 HALERYKKGKPRFLVLDLNOG 380
Db 451 --LKKFLSKPKILVFNIRKG 469
```

```
RESULT 9
DEGP_RICCN STANDARD; PRT; 508 AA.
AC Q92JAI;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR RC0166.
OS Rickettsia conorii.
CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

FT	SIGNAL	1	26	POTENTIAL.
FT	CHAIN	27	478	PROBABLE SERINE PROTEASE DO-LIKE.
FT	DOMAIN	116	254	CATALYTIC.
FT	DOMAIN	281	372	PDZ 1.
FT	DOMAIN	387	469	PDZ 2.
FT	ACT_SITE	133	133	CHARGE RELAY SYSTEM (POTENTIAL).
FT	ACT_SITE	163	163	CHARGE RELAY SYSTEM (POTENTIAL).
FT	ACT_SITE	238	238	CHARGE RELAY SYSTEM (POTENTIAL).
SEQ	SEQUENCE	478 AA;	51303 MW;	C044824F7EF4E98E CRC64;
Query Match 33.9%; Score 657.5; DB 1; Length 478;				
Best Local Similarity 39.4%; Pred. No. 1.1e-31;				
Matches 148; Conservative 82; Mismatches 121; Indels 25; Gaps				
QY	10	ALCGSVTIISKD-GYIVTNHVIDGADKIKVTIPGSKNKEYSATLVGTDSSDLAVIRI-TK 67		
DB	114	ALCGSVIINADKAYATNHHVVENANKIQVL-SDGRRYEASIIKGRSDIALQLKNA 172		
QY	68	DNLPTIKFSDNDISYGDILVFAIGNPFGVGSVTCQIVSALNKNKSGIGINSYENFIQTDDAS 127		
DB	173	KNLSAIKIAKSDTLRVGDVTVTAIGNPYGLGETVTSIIISALGRSGLENIEHYENFIQTDA 232		
QY	128	INPNSGGALIDSRGLGVINTAIIISKTGNGHCIGFAIPASNNVVKDPVTOLIKTGKIERGY 187		
DB	233	INRNSGGALVNLKGLIGINTAIIALAPDGGNGIGIIFAPGNMVKNLTEQMVFGQVKRGE 292		
QY	188	LGVLGDLDSGLDQ--SYNKEGAVVISYEKDSPAKKAGILVMDLTETVNGKKVKNTEL 245		
DB	293	LGIIIGELNSDLAHVMKINAKGAFVSQVLPSNSAFHAGIKAGDIIVSLNKKTIISSFAAL 352		
QY	246	RNLIGSMLPNQRVTLKVIKDKKERAPFTLTLAERKNPNKKTETISAOQAQGL-NGLOVED 304		
DB	353	RAEVGSILPVSTKMEIGFRN---GITKNVIVELKPSLKNVSYL-----GDIVTGIEGAD 403		
QY	305	LTQETKRSMLSDVDGVLVSQVNSPNSPAEQAGFROGNIITIEEVEVKSADFNHLEK 364		
DB	404	LSDCSLNGQK-----GVKIENIKLNTQASKIGFKKDDIIVEYNQKVINNLNDLNKILD- 456		
QY	365	YKCKPKRFVLVDLNOG 380		
DB	457	--SKP-NILVFSVKRG 469		
RESULT 11				
ID	DEQO_ECOLI	STANDARD;	PRT;	455 AA.
AC	P39099;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DE	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Protease deqO precursor (EC 3.4.21.-).			
GN	DEQO OR HWOA OR B3234.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12 / W3110;			
RA	Bass S., Gu Q., Goddard A.;			
RL	Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.			
RC	STRAIN-K12 / W3110;			
RA	MEDLINE-96165272; PubMed-8576051;			
RX	Waller P.R., Sauer R.T.;			
RT	"Characterization of deqO and degS, Escherichia coli genes encoding			
RT	homologs of the deqO protease.";			
RL	J. Bacteriol. 178:1146-1153(1996).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12 / MG1655;			
RX	MEDLINE-97426617; PubMed-9278503;			

```
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15661; AAC43992.1; -.
DR EMBL; U32495; AAC44005.1; -.
DR EMBL; U18997; AAA58036.1; -.
DR EMBL; AE000402; AAC76266.1; -.
DR MEROPS; S01.274; -.
DR SWISS-2DPAGE; P39099; COLI.
DR EcoGene; EG12612; degQ.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 455
FT DOMAIN 258 349
FT DOMAIN 355 447
FT ACT_SITE 109 109
FT ACT_SITE 139 139
FT ACT_SITE 214 214
FT ACT_SITE 455 455
SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;

Query Match 33.6%; Score 652; DB 1; Length 455;
Best Local Similarity 38.7%; Pred. No. 2.1e-31;
Matches 150; Conservative 76; Mismatches 130; Indels 32; Gaps 8;

QY 3 PKRMERALGSGVII-SKDGIVITNNHVIDGADKIKVTIPGSKNKEYSATVLGTDSESDIA 61
DB 84 PAQPF-GLGSGVIINASKGYVLTNNHVINQAQKISIQ-L-NDGREFAKLIGSDQSDIA 141
QY 62 VIRI-TKNLPIKSDSNDISVDGLVFAIGNPFGVSGESVTGIVSALNKSGIGINSVEN 120
DB 142 LLIQIKNPSKLTQIATADSKLVGDFAVAVGNPFGIGQTATSGIVSALGRSLNLEGLN 201
QY 121 FIQTASINPGSGALLDSRGLGVINTAIITSKTGGNHGIGFAIPSNMVKDVTQLIKT 180
DB 202 FIQTASINRNGSGALLNLNGELIGINTAILAPGGVSGIGFAIPSNMARTLAQLIDF 261
QY 181 KXIERGYLGVGLQDLSGLDQNSY--DNKEGAVVIVSEKDSAPKAKGILVWDLITVEVNGK 238
DB 262 GEIKRGLLGKIGTENSADIAKFNLDVQGFVSEVLPFGSGSAKAGVAGDIITSINGKP 321
QY 239 VKNTELNLNIGSMLPNORVTLKVRDKK--ERATFLIAERKNPNKKTETISAQNAQOQ 296
DB 322 LNSFAELRSRIATTEPGTKVGLGRNGKPLEVEVTLDTSTSSASAEMITPALEGA--- 378
QY 297 LNLQVEDLTQETKSRMLSDDDVQVLYSOVNNENSPAQAGPRGNIITKIEVEVKSVA 356
DB 379 -----TLDGQLKDGKGKGIKIDVYGVKSPAAQAGLQKDDVDIIVNRDRVNSIA 426
```

```
QY 357 DFNH-----ALEKYKGKPKRFLVL 375
DB 427 EMRKVLAAPAIIALQIVRGNESIYLLM 454

RESULT 12
DEGP_RICPR
ID DEGP_RICPR STANDARD; PRT; 513 AA.
AC O05942;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR Rpl24.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN [2]
RP SEQUENCE OF 161-513 FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=97419517; PubMed=9274032;
RA Andersson J.O., Andersson S.G.E.;
RT "Genomic rearrangements during evolution of the obligate intracellular
RT parasite Rickettsia prowazekii as inferred from an analysis of 52015
RL Microbiology 143:2783-2795(1997).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235270; CNA14593.1; -.
DR EMBL; Y11782; CAA72471.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001234; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 513
FT DOMAIN 124 289
FT DOMAIN 290 381
FT DOMAIN 418 502
FT ACT_SITE 139 139
FT ACT_SITE 169 169
FT ACT_SITE 247 247
FT ACT_SITE 513 513
SQ SEQUENCE 513 AA; 56309 MW; 4B7E9B7AB4079139 CRC64;

Query Match 32.3%; Score 625.5; DB 1; Length 513;
Best Local Similarity 37.7%; Pred. No. 8.8e-30;
Matches 146; Conservative 79; Mismatches 139; Indels 23; Gaps 9;
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:28:01 ; Search time 22 seconds
(without alignments)
1691.092 Million cell updates/sec

Title: US-09-895-913A-120
Perfect score: 1938
Sequence: 1 MIPKMERALSGVVISKD.....KPKRFLVLDLNOGYRIILVK 387

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1938	100.0	443	1 C64647	serine proteinase
2	1914	98.8	476	2 H71936	proteinase DO - He
3	937	48.3	472	2 F81329	serine proteinase
4	767	39.6	456	2 F82307	proteinase DO VC05
5	749.5	38.7	497	2 H71465	probable do serine
6	746.5	38.5	497	2 B81728	serine proteinase,
7	722.5	37.3	474	2 S45229	proteinase DO (EC
8	722.5	37.3	474	2 E85500	proteinase DO (EC
9	722.5	37.3	474	2 E90649	proteinase DO (EC
10	722.5	37.3	488	2 H86612	DO serine proteinase
11	722.5	37.3	488	2 G81528	serine proteinase,
12	722.5	37.3	488	2 G72011	do serine proteinase
13	717.5	37.0	475	2 AC0528	protease DO precu
14	714.5	36.9	475	1 S15337	heat shock protein
15	710	36.6	475	1 B70426	periplasmic serine
16	707.5	36.5	474	2 A13349	proteinase DO (EC
17	706.5	36.5	474	2 F83550	serine proteinase
18	703.5	36.3	474	1 I40059	htrA-like protein
19	698	36.0	457	2 AG0433	proteinase (EC 3.4
20	692.5	35.7	523	2 A97479	probable serine pr
21	692.5	35.7	523	2 A12696	serine proteinase
22	679.5	35.1	513	2 I40060	serine proteinase
23	679.5	35.1	513	2 A34118	proteinase DO (EC
24	679	35.0	466	1 A64113	heat shock protein
25	678.5	35.0	478	2 G84956	proteinase do prec
26	667.5	34.4	481	2 AH0410	global stress requ
27	663.5	34.2	511	2 F97720	periplasmic serine
28	655	33.8	455	2 AB0909	serine protease (E
29	653	33.7	514	2 A82581	periplasmic protei

ALIGNMENTS

RESULT 1

C64647

serine proteinase (EC 3.4.21.-) - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jun-2000

C:Accession: C64647

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R

son, J.D.; Kelley, J.M.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A61520; MUID:97394467; PMID:9252185

A:Accession: C64647

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-443 <TOM>

A:Cross-references: GB:AE000610; GB:AE000511; NID:g2314160; PID:AA08063.1; PID:g231

C:Genetics:

A:Start codon: GTG

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

Query Match 100.0% Score 1938; DB 1; Length 443;
Best Local Similarity 100.0%; Pred. No. 2.9e-111;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIPKMERALSGVVISKDGIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTDSDDL 60
|||||
Db 57 MIPKMERALSGVVISKDGIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTDSDDL 116
|||||

Qy 61 AVIRITKONLPTIKFSDNSDISVGLDVFAGNPFVGESVTOGIVSALNKSIGINSYEN 120
|||||

Db 117 AVIRITKONLPTIKFSDNSDISVGLDVFAGNPFVGESVTOGIVSALNKSIGINSYEN 176
|||||

Qy 121 FIQTASINPGNSGALLDSRGLVGINTAIISKTTGGNHGIGFAIPSNMVKDTVTLIKT 180
|||||

Db 177 FIQTASINPGNSGALLDSRGLVGINTAIISKTTGGNHGIGFAIPSNMVKDTVTLIKT 236
|||||

Qy 181 GRIERYGLGVGLQDLSGDLQNSYDNKEGAVISVEKDSPAKKAGILVWDLITEVNGKKVK 240
|||||

Db 237 GRIERYGLGVGLQDLSGDLQNSYDNKEGAVISVEKDSPAKKAGILVWDLITEVNGKKVK 296
|||||

Qy 241 NTNELNLIGSMQPNQRTLVKIVRDKKERAFTLTIAERKNPNKKTISAQNGAQQNLGL 300
|||||

Db 297 NTNELNLIGSMQPNQRTLVKIVRDKKERAFTLTIAERKNPNKKTISAQNGAQQNLGL 356
|||||

Qy 301 QVEDLTQETKRSRLSDDDVQGVLVSVQVNSPAEQAFRQGNIIITKIEVEVKSVDNFH 360
|||||

Db 357 QVEDLTQETKRSRLSDDDVQGVLVSVQVNSPAEQAFRQGNIIITKIEVEVKSVDNFH 416
|||||

Qy 361 ALEKYKGKPKRFLVLDLNOGYRIILVK 387

|||||
Db 417 ALEKYGKPRFLVLDLNOGYRIILVK 443

RESULT 2
H71936
Proteinase DO - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Sep-1999
C:Accession: H71936
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jlang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: H71936
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-476 <ARN>
A:Cross-references: GB:AE001474; GB:AE001439; NID:g4154929; PIDN:AA05980.1; PID:g415493
A:Experimental source: strain J99
C:Genetics:
A:Gene: htrA
C:Superfamily: Helicobacter serine proteinase

Query Match 98.8%; Score 1914; DB 2; Length 476;
Best Local Similarity 98.7%; Pred. No. 9.3e-110;
Matches 382; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MIPKMERALGSGVVISKDGYYVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESL 60
Db 90 MIPKMERALGSGVVISKDGYYVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESL 149
QY 61 AVIRITKDNLTPTKFSNDSNDISVGLVFAIGNPFGVESVTOGIVSALNKGSGINSYEN 120
Db 150 AVIRITKDNLTPTKFSNDSNDILVGLVFAIGNPFGVESVTOGIVSALNKGSGINSYEN 209
QY 121 FIQTASINPNSGGALIDSRGLGVINTAIISKTGNGHIGFAIPSNMVKDVTQTILKT 180
Db 210 FIQTASINPNSGGALIDSRGLGVINTAIISKTGNGHIGFAIPSNMVKDVTQTILKT 269
QY 181 GKTERGYLVGLQDLSGDLQNSYDNKEGAVIVSEKDSAPKAGILVWDLITEVNGKKVK 240
Db 270 GKTERGYLVGLQDLSGDLQNSYDNKEGAVIVSEKDSAPKAGILVWDLITEVNGKKVK 329
QY 241 NTNELRNLGSMPLPNORVTLKVRDKKERAFITLTLAERKNPNKKTISAQNGAQQGLNGL 300
Db 330 NTNELRNLGSMPLPNORVTLKVRDKKERAFITLTLAERKNPNKKTISAQNGVQGGNGL 389
QY 301 QVEDLTQETKRSMRLSDDVQGVLVSVQVNSPAEQAGFRQGNITIKIEEVEKSVADFNNH 360
Db 390 QVEDLTQETKRSMRLSDDVQGVLVSVQVNSPAEQAGFRQGNITIKIEEVEKSVADFNNH 449
QY 361 ALEKYGKPRFLVLDLNOGYRIILVK 387
Db 450 ALEKYGKPRFLVLDLNOGYRIILVK 476

RESULT 3
F81329
serine proteinase (proteinase DO) (EC 3.4.21.-) Cj1228c [imported] - Campylobacter jejuni
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: F81329
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrer
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: F81329
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-472 <PAR>
A:Cross-references: GB:ALJ139077; GB:ALJ11168; NID:g6968444; PIDN:CAB73482.1; PID:g696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: htrA; Cj1228c
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 48.3%; Score 937; DB 2; Length 472;
Best Local Similarity 51.3%; Pred. No. 5.5e-50;
Matches 195; Conservative 69; Mismatches 102; Indels 14; Gaps 5;

QY 10 ALGSGVVISKDGYYVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRITKDN 69
Db 102 ALGSGVVISKDGYYVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAVIRITKDN 161
QY 70 LPTIKFSDNSNDISVGLVFAIGNPFGVESVTOGIVSALNKGSGINSYENFQTQDASIN 129
Db 162 LSAITFTNSDLDMEGDDVVFALGNPFGVSVTSGIISALNKGSGINSYENFQTQDASIN 221
QY 130 PNSSGGALIDSRGLGVINTAIISKTGNGHIGFAIPSNMVKDVTQTILKTGRIERYGL 189
Db 222 PNSSGGALIDSRGLGVINSAILSRGSGNNGIGFAIPSNMVKDIAKKLIEKGIDRGFLG 281
QY 190 VGLQDLSGDLQNSYDNKEGAVIVSEKDSAPKAGILVWDLITEVNGKKVKNTNELRNLI 249
Db 282 VTILALQGDTKKAYKNQEGALITDVQKGSADAGLRGDLVTKVNDKVIKSPIDLKNIYI 341
QY 250 GSNLPMQVRLTKVIRD--KKERAFITLTLAERKNPNKKTISAQNGAQQGL--NGLQVEDLT 306
Db 342 GTLEIGOKTSLSYERDGENKNQASFIK--GEKENP-----KGVSDDLIDGLSLRNLD 391
QY 307 QETKRSMRLSDDVQGVLVSVQVNSPAEQAGFRQGNITIKIEEVEKSVADFNNHALEKVK 366
Db 392 PRKLDRLQIPKQVNGVLDVSVKESKSGSGFQEGDIIIGVGSEIKNLKDLQAL--KQV 450
QY 367 GPKRFLVLDLNOGYRIILV 386
Db 451 NKKEFTKVVYRNGFATLLV 470

RESULT 4
F82307
proteinase DO VC0566 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82307
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: F82307
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-456 <HEI>
A:Cross-references: GB:AE004142; GB:AE003852; NID:g654988; PIDN:AAF93734.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0566
A:Map position: 1
C:Superfamily: Helicobacter serine proteinase

Query Match 39.6%; Score 767; DB 2; Length 456;
Best Local Similarity 43.6%; Pred. No. 1.3e-39;
Matches 168; Conservative 83; Mismatches 112; Indels 22; Gaps 8;

QY 4 KERMERALGSGVVISKDGYYVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAV 62
Db 85 QERPFRGLGSGVVISKDGYYVTNNHVIDGADKIKVTIPGSKNKEYSATLVGTDSESLAV 143
QY 63 IRITK--DNLPTKFSNDSNDISVGLVFAIGNPFGVESVTOGIVSALNKGSGINSYENF 121

Db 144 LKLNKAKNLTIRIADSKLRVGDFAVAGNPFGLGQVTVSGIVSALGRSLNTEFNFE 203
Qy 122 IQTDAINPNSGGALIDSRGGLVGINTAIISKTGGNHGIGFAIPSNMVKDTVTOLIKTG 181
Db 204 IOTDAINSGSGGALVNLGELIGINTAILGPNNGVIGFAIPSNMKNLTDOLIEFG 263
Qy 182 KIERGYLVGLQDLSGDLQNS--YDNKEGAVVISVEKDSAPPAKAGILVMDLITEVNGKKV 239
Db 264 EVKRGMLGVQGEITSELADALGVESKGAFAVSQVVPDAAADKAGIKAGDIITSLNGKKI 323
Qy 240 KNTNELNLIGSMLPNQVTLKVRDKKERAFITLAEKKNPNKKTISAQNGAQQLNG 299
Db 324 DTFSELRAKAVATLGAGKTITIGVLDRGNQINDVTLGSEQNAKTK----AESLHQG-LSG 378
Qy 300 LQVEDLTQETKRSMRLSDVQGLVSVQVNSPBAQAGFRQGNITTKIEEVEKVSADFEN 359
Db 379 AELSNTD-----SDPIQGVKTEVQKGSAAESYQLQDKDIIIGVNRKRVKNAIELR 430
Qy 360 HALEYKKGKPRKRFVLVDLNOQYRII 384
Db 431 AIME-----KSPNIALNLTQRGERTL 451

RESULT 5
H71465
A:Query Match 38.7%; Score 749.5; DB 2; Length 497;
Best Local Similarity 41.3%; Pred. No. 1.7e-38;
Matches 161; Conservative 84; Mismatches 126; Indels 19; Gaps 7;
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: H71465
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
A:Reference number: A71570; MUID:95000809; PMID:9784136
A:Accession: H71465
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-497 <R>
A:Cross-references: GB:AB001355; GB:AB001273; NID:g3329292; PIDN:AAC68420.1; PID:g332929
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: htrA
C:Superfamily: Helicobacter serine proteinase

Query Match 38.7%; Score 749.5; DB 2; Length 497;
Best Local Similarity 41.3%; Pred. No. 1.7e-38;
Matches 161; Conservative 84; Mismatches 126; Indels 19; Gaps 7;
Qy 3 PKERMALGSGVIISKDGYIVTNHVDGADKIKVTIPGSNKEYSATLVGTDSSEDLAV 62
Db 119 PQQR-DAVGRGTGFIIVSEDDGYVVTNHVVEDAGKIHVTLHDGQK-YTAKIIGLDPKTDLAV 176
Qy 63 IRIKDNLPITKFSNDSIDVGLVFAIGNPFGVGVSTQGVISALNSKSGIGINSYENFI 122
Db 177 IKIQAEKLPFTLFGNSDQGLGDMNAIAGNPFGLQATVTVGVISAKGRNQLHIVDFEDI 236
Qy 123 QTDASINPNSGGALIDSRGGLVGINTAIISKTGGNHGIGFAIPSNMVKDTVTOLIKTG 182
Db 237 QTDRAINPNSGGALVNLGELIGINTAILGPNNGVIGFAIPSNMKNLTDOLIEFG 296
Qy 182 KIERGYLVGLQDLSGDLQNSY--DNKEGAVVISVEKDSAPPAKAGILVMDLITEVNGKKV 240
Db 297 VTRGFLGVTLQPIDSELATCYKLEKVGALVTVVKGSPAEKAGLRQEDVIVAYNGKEVE 356
Qy 241 NTNELNLIGSMLPNQVTLKVRDKKERAFITLAEKKNPNKKTISAQNGAQ-QLNG 299
Db 357 SLSALRNAISLMPGTRVVLKVRREGKTIEIPVTVTQ-----IPTEDGVSAQKMG 407
Qy 300 LQVEDLTQETKRSMRLSDVQGLVSVQVNSPBAQAGFRQGNITTKIEEVEKVSADFEN 359
Db 408 VRVQNITPEICKKGLADTIGILVAVVAGSPASAGVAPGQLILAVNRQRVASVEELN 467
Qy 360 HALEYKKGKPRKRFVLVDLNOQ--YRIILVK 387

Db 468 QVLKNSKGEN---VLLMVSQGDVVRFVLK 494
RESULT 6
B81728
A:Query Match 38.5%; Score 746.5; DB 2; Length 497;
Best Local Similarity 41.3%; Pred. No. 2.7e-38;
Matches 161; Conservative 84; Mismatches 126; Indels 19; Gaps 7;
Qy 3 PKERMALGSGVIISKDGYIVTNHVDGADKIKVTIPGSNKEYSATLVGTDSSEDLAV 62
Db 119 PQQR-DAVGRGTGFIIVSEDDGYVVTNHVVEDAGKIHVTLHDGQK-YTAKIIGLDPKTDLAV 176
Qy 63 IRIKDNLPITKFSNDSIDVGLVFAIGNPFGVGVSTQGVISALNSKSGIGINSYENFI 122
Db 177 IKIQAEKLPFTLFGNSDQGLGDMNAIAGNPFGLQATVTVGVISAKGRNQLHIVDFEDI 236
Qy 123 QTDASINPNSGGALIDSRGGLVGINTAIISKTGGNHGIGFAIPSNMVKDTVTOLIKTG 182
Db 237 QTDRAINPNSGGALVNLGELIGINTAILGPNNGVIGFAIPSNMKNLTDOLIEFG 296
Qy 182 KIERGYLVGLQDLSGDLQNSY--DNKEGAVVISVEKDSAPPAKAGILVMDLITEVNGKKV 240
Db 297 VTRGFLGVTLQPIDSELATCYKLEKVGALVTVVKGSPAEKAGLRQEDVIVAYNGKEVE 356
Qy 241 NTNELNLIGSMLPNQVTLKVRDKKERAFITLAEKKNPNKKTISAQNGAQ-OLNG 299
Db 357 SLSALRNAISLMPGTRVVLKVRREGKTIEIPVTVTQ-----IPAEDGVSAQKMG 407
Qy 300 LQVEDLTQETKRSMRLSDVQGLVSVQVNSPBAQAGFRQGNITTKIEEVEKVSADFEN 359
Db 408 VRVQNITPEICKKGLADTIGILVAVVAGSPASAGVAPGQLILAVNRQRVASVEELN 467
Qy 360 HALEYKKGKPRKRFVLVDLNOQ--YRIILVK 387
Db 468 QVLKNAKGEN---VLLMVSQGEVIRFVLK 494

RESULT 7
S45229
A:Query Match 38.5%; Score 746.5; DB 2; Length 497;
Best Local Similarity 41.3%; Pred. No. 2.7e-38;
Matches 161; Conservative 84; Mismatches 126; Indels 19; Gaps 7;
Qy 3 PKERMALGSGVIISKDGYIVTNHVDGADKIKVTIPGSNKEYSATLVGTDSSEDLAV 62
Db 119 PQQR-DAVGRGTGFIIVSEDDGYVVTNHVVEDAGKIHVTLHDGQK-YTAKIIGLDPKTDLAV 176
Qy 63 IRIKDNLPITKFSNDSIDVGLVFAIGNPFGVGVSTQGVISALNSKSGIGINSYENFI 122
Db 177 IKIQAEKLPFTLFGNSDQGLGDMNAIAGNPFGLQATVTVGVISAKGRNQLHIVDFEDI 236
Qy 123 QTDASINPNSGGALIDSRGGLVGINTAIISKTGGNHGIGFAIPSNMVKDTVTOLIKTG 182
Db 237 QTDRAINPNSGGALVNLGELIGINTAILGPNNGVIGFAIPSNMKNLTDOLIEFG 296
Qy 182 KIERGYLVGLQDLSGDLQNSY--DNKEGAVVISVEKDSAPPAKAGILVMDLITEVNGKKV 240
Db 297 VTRGFLGVTLQPIDSELATCYKLEKVGALVTVVKGSPAEKAGLRQEDVIVAYNGKEVE 356
Qy 241 NTNELNLIGSMLPNQVTLKVRDKKERAFITLAEKKNPNKKTISAQNGAQ-OLNG 299
Db 357 SLSALRNAISLMPGTRVVLKVRREGKTIEIPVTVTQ-----IPAEDGVSAQKMG 407
Qy 300 LQVEDLTQETKRSMRLSDVQGLVSVQVNSPBAQAGFRQGNITTKIEEVEKVSADFEN 359
Db 408 VRVQNITPEICKKGLADTIGILVAVVAGSPASAGVAPGQLILAVNRQRVASVEELN 467
Qy 360 HALEYKKGKPRKRFVLVDLNOQ--YRIILVK 387
Db 468 QVLKNAKGEN---VLLMVSQGEVIRFVLK 494

proteinase DO (EC 3.4.21.1) precursor / heat shock protein htrA - Escherichia coli (s
C:Species: Escherichia coli
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Mar-2002
C:Accession: S45229; A:4740; S01899; B35993
R:Fujita, N.
submitted to the EMBL data Library, January 1994
A:Reference number: S45181
A:Accession: S45229
A:Molecule type: DNA
A:Residues: 1-474 <F>
A:Cross-references: EMBL:D26562; NID:q473770; PIDN:BAA05608.1; PID:q473819
A:Experimental source: strain K-12, substrain W3110
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: A64740

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-474 <BLAT>

A:Cross-references: GB:AE000125; GB:U00096; NID:gl1786348; PIDN:AAC73272.1; PID:gl1786356;

A:Experimental source: strain K-12, substrain MGL1655

R:Lipinska, B.; Sharma, S.; Georgopoulos, C.

Nucleic Acids Res. 16, 10053-10067, 1988

A:Title: Sequence analysis and regulation of the htrA gene of *Escherichia coli*: a sigma

A:Reference number: S01899; MUID:89057448; PMID:3057437

A:Accession: S01899

A:Molecule type: DNA

A:Residues: 1-9,'R',11-191,'G',193-466,'R'HPVNAVLSNPLKTKGRSPYNL' <LIP>

A:Cross-references: EMBL:X12457; NID:941760; PIDN:CAA30997.1; PID:g41761

A:Experimental source: strain K-12

R:Wurgler, S.M.; Richardson, C.C.

Proc. Natl. Acad. Sci. U.S.A. 87, 2740-2744, 1990

A:Title: Structure and regulation of the gene for dGMP triphosphohydrolase from *Escherichia coli*

A:Reference number: A35993; MUID:90207273; PMID:2157212

A:Accession: B35993

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-9,'R',11-16 <WUR>

A:Cross-references: GB:M31772; NID:gl45733; PIDN:AAA23680.1; PID:gl45736

C:Genetics:

A:Gene: htrA

A:Map position: 4 min

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-474/Product: heat shock protein htrA #status predicted <MAT>

Query Match 37.3%; Score 722.5; DB 2; Length 474;

Best Local Similarity 43.5%; Pred. No. 7.3e-37;

Matches 163; Conservative 74; Mismatches 113; Indels 25; Gaps 9;

QY 10 ALGSGVLIISKD-GYIVTNHVIDGADKIKVTIPGSNKYSATLVGDSGLAVIRI-TK 67

Db 112 ALGSGVLIIDAKGYYVTVNNHVVVDNATVIKVL-SDGRKFDKMKVGDPRSDIALIQINP 170

QY 68 DNLPTIKFSDNSISVGLVFAIGNPFGVSGESVTQIGVSALNKSIGINSYENFIOTDAS 127

Db 171 KNLTAIKMADSDALRVGDYTVAGNPFGLGETVTSIGVSALGRSGLNAENYENFIOTDAA 230

QY 128 INPNSGGALIDSRGLVGINTAIISKTGNHIGIGFPAIPSNMVKDVTVOLIKTKIERGY 187

Db 231 INRNSGGALVNLNGELIGINTAILAPDGNIGIGFPAIPSNMVKLTQSMVEYGVKRG 290

QY 188 LGVGLQDLSGLQNS--YDNKEGAVVISVEKDSPAKAGILVWDLITEVNGKVKVKNTEL 245

Db 291 LGTMGTSELSELAKAMKVDQAGFVSQVLPNSAAKAGIKAGDVITSLNGKRPISFAAL 350

QY 246 RNLIGSLNPNQRTLVKIVIRDKKRAFTTLAERKNPNKKTETISAQNGAQLNGLOVEDL 305

Db 351 RAQVGTMPVGSKITLGLLRDQKQVNNLEL-QOSSQNVQDSSSIFNGIEG----- 399

QY 306 TQTKSRMLSDVQGVLSQVNSNENSPAQAGPRQGNITKIEVEVKSVADFNHALEYK 365

Db 400 ---AEMSNKGD--QGVVNVNKTGTTPAAQIGLKGDDVIIGANQQAVKNTAELRKVLDD-- 452

QY 366 KGKPKRFLVDLNOQ 380

Db 453 -SKPS-VLALNTIQRG 465

RESULT 8

E85500

proteinase DO (EC 3.4.21.-) precursor / heat shock protein htrA - *Escherichia coli* (stra

C:Species: *Escherichia coli*

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002

C:Accession: E85500

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E85500

A:Molecule type: DNA

A:Residues: 1-474 <STO>

A:Cross-references: GB:AE005174; NID:gl12512885; PIDN:AA054465.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: htrA

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

Query Match 37.3%; Score 722.5; DB 2; Length 474;

Best Local Similarity 43.5%; Pred. No. 7.3e-37;

Matches 163; Conservative 74; Mismatches 113; Indels 25; Gaps 9;

QY 10 ALGSGVLIISKD-GYIVTNHVIDGADKIKVTIPGSNKYSATLVGDSGLAVIRI-TK 67

Db 112 ALGSGVLIIDAKGYYVTVNNHVVVDNATVIKVL-SDGRKFDKMKVGDPRSDIALIQINP 170

QY 68 DNLPTIKFSDNSISVGLVFAIGNPFGVSGESVTQIGVSALNKSIGINSYENFIOTDAS 127

Db 171 KNLTAIKMADSDALRVGDYTVAGNPFGLGETVTSIGVSALGRSGLNAENYENFIOTDAA 230

QY 128 INPNSGGALIDSRGLVGINTAIISKTGNHIGIGFPAIPSNMVKDVTVOLIKTKIERGY 187

Db 231 INRNSGGALVNLNGELIGINTAILAPDGNIGIGFPAIPSNMVKLTQSMVEYGVKRG 290

QY 188 LGVGLQDLSGLQNS--YDNKEGAVVISVEKDSPAKAGILVWDLITEVNGKVKVKNTEL 245

Db 291 LGTMGTSELSELAKAMKVDQAGFVSQVLPNSAAKAGIKAGDVITSLNGKRPISFAAL 350

QY 246 RNLIGSLNPNQRTLVKIVIRDKKRAFTTLAERKNPNKKTETISAQNGAQLNGLOVEDL 305

Db 351 RAQVGTMPVGSKITLGLLRDQKQVNNLEL-QOSSQNVQDSSSIFNGIEG----- 399

QY 306 TQTKSRMLSDVQGVLSQVNSNENSPAQAGPRQGNITKIEVEVKSVADFNHALEYK 365

Db 400 ---AEMSNKGD--QGVVNVNKTGTTPAAQIGLKGDDVIIGANQQAVKNTAELRKVLDD-- 452

QY 366 KGKPKRFLVDLNOQ 380

Db 453 -SKPS-VLALNTIQRG 465

RESULT 9

E90649

proteinase DO (EC 3.4.21.-) precursor / heat shock protein Ecs0165 [Imported] - *Esche*

C:Species: *Escherichia coli*

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002

C:Accession: E90649

R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: E90649

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-474 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA033588.1; PID:gl3359621; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ecs0165

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

Query Match 37.3%; Score 722.5; DB 2; Length 474;
Best Local Similarity 43.5%; Pred. No. 7.7e-37;
Matches 163; Conservative 74; Mismatches 113; Indels 25; Gaps 9;

Qy 10 ALGSGVIRSKD-GYIVNNHVIDGADKIKVTIPGSNKEYSATVGTDSSEDLAVIRI-TK 67
Dy 112 ALGSGVIRSKD-GYIVNNHVIDGADKIKVTIPGSNKEYSATVGTDSSEDLAVIRI-TK 67
Qy 68 DNLPTIKFSDNDISVGLDFAIGNPFGVSGVTSALNKGSGIGINSYENFIQTDAS 127
Dy 171 KNLTAIRKWDSDALRVGDYTVAINPFGVSGVTSALNKGSGIGINSYENFIQTDAS 127
Qy 128 INPNSGGALIDSRGLVGINTAISKTTGGNHGIFAIIPSNMVKDTVTQLIKTKIERGY 187
Dy 231 INRNSGGALVNLGELIGINTAILAPDGGNIGIGIFAIIPSNMVKDTVTQLIKTKIERGY 187
Qy 188 LGVGLQDLSGLDQNS--YDNKEGAVTISVEKDSAPKAGILVWDLITVEVNGKKVKNTEL 245
Dy 291 LGVGLQDLSGLDQNS--YDNKEGAVTISVEKDSAPKAGILVWDLITVEVNGKKVKNTEL 245
Qy 246 RNLIGSMPLPNORVTLKVRDKKERAFTLTTLAERKNPNKKTETISAQNGQGLNGLOVEDL 305
Dy 351 RAQVGTMPVSGKLILGLLRDQGVNNLEL-QQSSQNOVDSSSIFNGIEG----- 399
Qy 306 TQETKRSMLSDDDVQGVLSQVNSPAAEQAGFRQGNITIKIEVEVKSVADEFNHALEKY 365
Dy 400 ---AEMSNKRGD--QGVVNNVKTGTAAQIGLKGVDVIGANQAVKNIAELRKVLD-- 452
Qy 366 KGKPKRFLVLDLNOG 380
Dy 453 -SKPS-VLALNIQRG 465

RESULT 10
H86612
Do serine proteinase [imported] - Chlamydomydia pneumoniae (strain J138)
C:Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: H86612
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: H86612
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <STO>
A:Cross-references: GB:BA000008; MID:98979352; PIDN:BAA99186.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: htrA
C:Superfamily: Helicobacter serine proteinase

Query Match 37.3%; Score 722.5; DB 2; Length 488;
Best Local Similarity 40.0%; Pred. No. 7.7e-37;
Matches 159; Conservative 82; Mismatches 134; Indels 21; Gaps 7;

Qy 2 IPKER-----MERALGSGVIRSKDGYIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTD 56
Dy 103 LPSOREKPSQKAEVRGTFIVSPDGYIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTD 56
Qy 57 ESDLAVIRITKDNLPFTIKFSDNDISVGLDFAIGNPFGVSGVTSALNKGSGIGIN 116
Dy 162 KTDLAVIKIKSONLPYLSFGNSDHLKVGDAIAGNPFGLQATVTVGVSAGRNQHLIA 221
Qy 117 SYENFIQTDASINPNSGGALIDSRGLVGINTAISKTTGGNHGIFAIIPSNMVKDTVTQ 176
Dy 222 DFEDFIQTDASINPNSGGALIDSRGLVGINTAISKTTGGNHGIFAIIPSNMVKDTVTQ 176
Qy 177 LIKTGKIERGYLVGLQDLSGLDQNSY--DNKEGAVTISVEKDSAPKAGILVWDLITEV 234
Dy 282 LIQDQVTRGFLVGLTLPQIDALACYKLEKVGALVTDVVKGSPADKAGLKQEDVIAT 341
Qy 235 NGKKVKTINELNRLIGSMPLPNORVTLKVRDKKERAFTLTTLAERKNPNKKTETISAQNG 294
Dy 342 NGKEVDSLSMFRNVAFLMNPOTRIVLKVVRGKVEIPVTSQAP---KEDGMSAL---- 394
Qy 295 GQNLGLQVEDITLQETKRSMLSDDDVQGVLSQVNSPAAEQAGFRQGNITIKIEVEVKS 354
Dy 395 -QRYGIRVQNTPTETAKKLGIAPETKGLIISVEPFGSVAASGIAPIGOLILAVNRQKRVSS 453
Qy 355 VADFNHALEKYGKPKRFLVLDLNOG--YRILVK 387
Dy 454 IEDLNRTL---KTSNNENILMWSSQGDVIRFIALK 485

RESULT 12
G72011
Do serine proteinase - Chlamydomydia pneumoniae (strain CWL029)
C:Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: G72011

Qy 235 NGKKVKTINELNRLIGSMPLPNORVTLKVRDKKERAFTLTTLAERKNPNKKTETISAQNG 294
Dy 342 NGKEVDSLSMFRNVAFLMNPOTRIVLKVVRGKVEIPVTSQAP---KEDGMSAL---- 394
Qy 295 GQNLGLQVEDITLQETKRSMLSDDDVQGVLSQVNSPAAEQAGFRQGNITIKIEVEVKS 354
Dy 395 -QRYGIRVQNTPTETAKKLGIAPETKGLIISVEPFGSVAASGIAPIGOLILAVNRQKRVSS 453
Qy 355 VADFNHALEKYGKPKRFLVLDLNOG--YRILVK 387
Dy 454 IEDLNRTL---KTSNNENILMWSSQGDVIRFIALK 485

RESULT 11
G81528
serine proteinase, HtrA/DegQ/DegS family CP0877 [imported] - Chlamydomydia pneumoniae
C:Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 21-Jul-2000
C:Accession: G81528
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: G81528
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <REA>
A:Cross-references: GB:AE002246; GB:AE002161; NID:97189785; PIDN:AAF38665.1; PID:9718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0877
C:Superfamily: Helicobacter serine proteinase

Query Match 37.3%; Score 722.5; DB 2; Length 488;
Best Local Similarity 40.0%; Pred. No. 7.7e-37;
Matches 159; Conservative 82; Mismatches 134; Indels 21; Gaps 7;

Qy 2 IPKER-----MERALGSGVIRSKDGYIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTD 56
Dy 103 LPSOREKPSQKAEVRGTFIVSPDGYIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTD 56
Qy 57 ESDLAVIRITKDNLPFTIKFSDNDISVGLDFAIGNPFGVSGVTSALNKGSGIGIN 116
Dy 162 KTDLAVIKIKSONLPYLSFGNSDHLKVGDAIAGNPFGLQATVTVGVSAGRNQHLIA 221
Qy 117 SYENFIQTDASINPNSGGALIDSRGLVGINTAISKTTGGNHGIFAIIPSNMVKDTVTQ 176
Dy 222 DFEDFIQTDASINPNSGGALIDSRGLVGINTAISKTTGGNHGIFAIIPSNMVKDTVTQ 176
Qy 177 LIKTGKIERGYLVGLQDLSGLDQNSY--DNKEGAVTISVEKDSAPKAGILVWDLITEV 234
Dy 282 LIQDQVTRGFLVGLTLPQIDALACYKLEKVGALVTDVVKGSPADKAGLKQEDVIAT 341
Qy 235 NGKKVKTINELNRLIGSMPLPNORVTLKVRDKKERAFTLTTLAERKNPNKKTETISAQNG 294
Dy 342 NGKEVDSLSMFRNVAFLMNPOTRIVLKVVRGKVEIPVTSQAP---KEDGMSAL---- 394
Qy 295 GQNLGLQVEDITLQETKRSMLSDDDVQGVLSQVNSPAAEQAGFRQGNITIKIEVEVKS 354
Dy 395 -QRYGIRVQNTPTETAKKLGIAPETKGLIISVEPFGSVAASGIAPIGOLILAVNRQKRVSS 453
Qy 355 VADFNHALEKYGKPKRFLVLDLNOG--YRILVK 387
Dy 454 IEDLNRTL---KTSNNENILMWSSQGDVIRFIALK 485

RESULT 12
G72011
Do serine proteinase - Chlamydomydia pneumoniae (strain CWL029)
C:Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: G72011

R:Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of *Clostridium botulinum* and *C. trachomatis*.
A:Reference number: AT7000; MUID:99206606; PMID:10192388
A:Accession: G72011
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <ARN>
A:Cross-references: GB:AE001678; GB:AE001363; NID:g4377301; PIDN:AA19116.1; PID:g437730
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: htrA
C:Superfamily: Helicobacter serine proteinase

Query Match 37.3%; Score 722.5; DB 2; Length 488;
Best Local Similarity 40.0%; Pred. No. 7.7e-37;
Matches 158; Conservative 82; Mismatches 134; Indels 21; Gaps 7;

QY 2 IPKER-----MERALSGVILSKDGIIVTNHVIDGADKIKVTIPGSKNKEYSATLVGTDSEDLAVIRI-TK 56
DB 103 LPSQREKQSKAVRGTFGLVSPDGIIVTNHVIDGADKIKVTIPGSKNKEYSATLVGTDSEDLAVIRI-TK 56
QY 57 ESDLAVIRITKDNLPKIFSDSDISVGDVFAIGNPFGVSGESVTQGIIVSALNKSIGIN 116
DB 162 KDLAVIKIKSNLPYLSFGNSDHLKVGDAIGAIFGLQATVTVGVISAKGRNQLHIA 221
QY 117 SYENFIQTDAINPGNSGALIDSRGLVGINTAISKTGNGHIGFAIPSNMVKDVTQ 176
DB 222 DFEDFIOTDAANPGNSGALLNIDQVIGVNTAIVSGSGYIGIGFAIPSLMANNRIIDQ 281
QY 177 LKTKGIERGYLGVLQDLSGLQNSY--DNKEGAVVISVEKDSPPAKKAGILVMDLITEV 234
DB 282 LTRDQGVTRGFLVQLPDAELACYLEKVGALVTDVWKGSPADKAGLQEDVIY 341
QY 235 NGKKVKNTELNLGSMPLNORVTLKVRDKKERAFVTLAERKNPNKKTETISAQNG 294
DB 342 NGKEVDSLMFNRAVSLAMPDTRIVLKVVRGKVIPIVTVSQAP---KEDGMSAL---- 394
QY 295 GQLNGQLVEDLTOETKRMSRLSDVGVLSQVNSNPAEQAGFRGNIITKIEEVEVK 354
DB 395 -QRVGIRVQNLTPETAKKLGIAPEKIGILISVEPSVAASSGAPQLLAVNRQKVS 453
QY 355 VADFNLAEKYGKPKRFLVLDLNOG--YRILVK 387
DB 454 IEDLNRTL---KDSNNENILLMVSQGVIRFALK 485

RESULT 13
AC0528
protease DO precursor, heat shock protein HtrA [imported] - *Salmonella enterica* subsp. e
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A:Note: this species has also been called *Salmonella typhi*
Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
Accession: AC0528
Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Ouaill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AC0528
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-475 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01363.1; PID:g16501490; GSPDB:GN00176
C:Genetics:
A:Gene: STY0231
C:Superfamily: Helicobacter serine proteinase

Query Match 37.0%; Score 717.5; DB 2; Length 475;
Best Local Similarity 43.2%; Pred. No. 1.5e-36;
Matches 162; Conservative 74; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVII-SKGGYIVTNHVIDGADKIKVTIPGSKNKEYSATLVGTDSEDLAVIRI-TK 67
DB 113 ALGSGVIIDAAKGYVTVTNHVIDNASVIKQL--SDGRKFDKAVVGDPRSDIALIQINP 171
QY 68 DNLPTIKFSDSDISVGDVFAIGNPFGVSGESVTQGIIVSALNKSIGINSENFITDAS 127
DB 172 KNLTAIKLADSDALRVGDYTVAINPFGVSGESVTQGIIVSALGRSGLNVENYFIQTDA 231
QY 128 INPGNSGALIDSRGLVGINTAISKTGNGHIGFAIPSNMVKDVTOLIKTKGIERGY 187
DB 232 INRNSGGALVNLNGELIGINTAILAPDGNIGIGFAIPSNMVKNTLSQWVEGVQVRGE 291
QY 188 LGVQLDLSGLQNS--YDNKEGAVVISVEKDSPPAKKAGILVMDLITEVNGKKVKNTEL 245
DB 292 LGIMGTSELSELAAMKVDQAGAFVSQVMPNNSAAKAGIKAGDVITSLNGKPISSFAAL 351
QY 246 RNLIGSMPLNORVTLKVRDKKERAFVTLAERKNPNKKTETISAQNGAQLNGLOVEDL 305
DB 352 RAQVGTMPPVGSKISLGLLREGK--AITVNL-----ELQSSSQSDSSSTIFSG 397
QY 306 TQETKRMSRLSDVGVLSQVNSNPAEQAGFRGNIITKIEEVEVKSVADFNHALEKY 365
DB 398 IGEAENSNGQD--KGVVSVKANSPPAAGIGLKKGDVIGANQOPVKVIAELRKILD-- 453
QY 366 KGKPKRFLVLDLNOG 380
DB 454 -SKPS-VLALNTIQRG 466

RESULT 14
S15337
heat shock protein htrA - *Salmonella typhimurium*
C:Species: *Salmonella typhimurium*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S15337; S21327
R:Johnson, K.; Charles, I.; Dougan, G.; Pickard, D.; O'Gaora, P.; Costa, G.; Ali, T.;
Mel. Microbiol. 5, 401-407, 1991
A:Title: The role of a stress-response protein in *Salmonella typhimurium* virulence.
A:Reference number: S15337; MUID:91251770; PMID:1645840
A:Accession: S15337
A:Molecule type: DNA
A:Residues: 1-475 <COS>
A:Cross-references: EMBL:X54548; NID:g47929; PID:g47930
C:Genetics:
A:Gene: htrA
C:Superfamily: Helicobacter serine proteinase

Query Match 36.9%; Score 714.5; DB 1; Length 475;
Best Local Similarity 42.9%; Pred. No. 2.3e-36;
Matches 161; Conservative 75; Mismatches 114; Indels 25; Gaps 9;

QY 10 ALGSGVII-SKGGYIVTNHVIDGADKIKVTIPGSKNKEYSATLVGTDSEDLAVIRI-TK 67
DB 113 ALGSGVIIDAAKGYVTVTNHVIDNASVIKQL--SDGRKFDKAVVGDPRSDIALIQINP 171
QY 68 DNLPTIKFSDSDISVGDVFAIGNPFGVSGESVTQGIIVSALNKSIGINSENFITDAS 127
DB 172 KNLTAIKLADSDALRVGDYTVAINPFGVSGESVTQGIIVSALGRSGLNVENYFIQTDA 231
QY 128 INPGNSGALIDSRGLVGINTAISKTGNGHIGFAIPSNMVKDVTOLIKTKGIERGY 187
DB 232 INRNSGGALVNLNGELIGINTAILAPDGNIGIGFAIPSNMVKNTLSQWVEGVQVRGE 291
QY 188 LGVQLDLSGLQNS--YDNKEGAVVISVEKDSPPAKKAGILVMDLITEVNGKKVKNTEL 245
DB 292 LGIMGTSELSELAAMKVDQAGAFVSQVMPNNSAAKAGIKAGDVITSLNGKPISSFAAL 351
QY 246 RNLIGSMPLNORVTLKVRDKKERAFVTLAERKNPNKKTETISAQNGAQLNGLOVEDL 305
DB 352 RAQVGTMPPVGSKISLGLLREGK--AITVNL-----ELQSSSQSDSSSTIFSG 397
QY 306 TQETKRMSRLSDVGVLSQVNSNPAEQAGFRGNIITKIEEVEVKSVADFNHALEKY 365

Db 398 IEAEMSNKQD--KGVVSVSKANSPAAIGLKKGDVLIIGANQQPVKNAIELRKILD-- 453
 QY 366 KGKPKRFLVLDLNOG 380
 Db 454 -SKPS-VLALNIORG 466

RESULT 15

B70426
 periplasmic serine proteinase (EC 3.4.21.-) - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: B70426
 R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oviatt, R.

Mature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: B70426
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-453 <AQF>

A:Cross-references: GB:AE000741; GB:AE000657; NID:g2983841; PID:AAC07399.1; PID:g298384
 A:Experimental source: strain VF5

C:Genetics:
 A:Gene: htrA
 C:Superfamily: Helicobacter serine proteinase
 C:Keywords: hydrolase; serine proteinase

Query Match 36.6%; Score 710; DB 1; Length 453;
 Best Local Similarity 41.2%; Pred. No. 4e-36;
 Matches 163; Conservative 81; Mismatches 126; Indels 26; Gaps 9;

QY 3 PRKMERALGSGVLIISKD-----YIVNNHVIGADKIKVILPGSNKEYSATLVGTDS 57

Db 67 PFTKRERSLGSGVIVKYVDEKVKVYILTNNAHVKNGVRIIVKL-DRHTEKKGGEIVGIDTK 125

QY 58 SDLAVIRITKDNLPIT-----KFSDSNDISVGDVLFVFAIGNPFGVGSVTQGITVSALNKS 112

Db 126 TDIAVVKISIRGINDIEDRIAKLGSDNLKVGQIVFAIGNPYGLERTVTMGVISALRRS- 184

QY 113 IGINSYENFIQTASINPGSGALIDSRGLVGINTALISKTGGNHGIGFAIPSNMVKD 172

Db 185 IGITQYESFIQTDAAINPGSGPLINVEGIVGINTAIIA---GAQGLGFAIPINLAKW 241

QY 173 TVTQLIKTKIERGYLGVLQDLGSLQNSYDNKEGANVISVEKDSPAKKAGILVWDLIT 232

Db 242 VMEQIIEHGKVIKRGWLVGVIQDITPDISEALGIGKGLVVAQVVPVGPSPADKAGLVGDVIV 301

QY 233 EVNGKKVKNTNELNLIGSMLPNQVTLAVIRDKKERAFETLAEKKNPKKETISAQNG 292

Db 302 EVNGKKIEDARDLQFTIMKKPGTKAVLVIRNGKEKEITVIIGQ-----YPEGVSRECK 357

QY 293 AQCQLNGLOVEDLTQETKRSMLSDVQGLVSVQNNENSPAQAGFROGNIITKIEVEV 352

Db 358 ATPENGLLLRLDLTLKEQEGV---YGVIVEGIPDSPAEGYSLQPGDIIILKYNRPV 414

QY 353 KSVADPNHALEYK--GKPKRFLVLDLNOGYRIILV 386

Db 415 RSVREFEYIINRLKEMGRSKALL--VRRGDRNIFI 448

Search completed: November 18, 2002, 11:30:06
 Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:33:31 ; Search time 22 Seconds
(without alignments)
1691.092 Million cell updates/sec

Title: US-09-895-913A-120
Perfect score: 387
Sequence: 1 MIPKMERALGSGVIISKD.....KPKRFLVLDLNOGYRIILVK 387

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR.73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	387	100.0	443	1 C64647	serine proteinase
2	24.8	96	476	2 H71936	proteinase DO - He
3	20	5.2	472	2 F81329	serine proteinase
4	18	4.7	481	2 D82826	heat shock protein
5	17	4.4	468	2 A12811	serine proteinase
6	17	4.4	495	2 B97590	htrA protein homol
7	14	3.6	348	2 H96956	serine protease Do
8	12	3.1	424	2 D89957	hypothetical prote
9	12	3.1	466	1 A64113	heat shock protein
10	12	3.1	474	1 I40059	htrA-like protein
11	12	3.1	474	2 S45229	proteinase DO (EC
12	12	3.1	474	2 A13349	proteinase DO (EC
13	12	3.1	474	2 E85500	proteinase DO (EC
14	12	3.1	474	2 E90649	proteinase DO (EC
15	12	3.1	475	1 S15337	heat shock protein
16	12	3.1	475	2 AC0528	protease DO precu
17	12	3.1	481	2 AH0410	global stress requ
18	11	2.8	408	2 H86891	exported serine pr
19	11	2.8	455	2 JC6051	trypsin-like prote
20	11	2.8	455	2 C91142	serine endoprotein
21	11	2.8	455	2 F85987	serine endoprotein
22	11	2.8	455	2 AB0909	serine protease (E
23	11	2.8	457	2 AG0433	proteinase (EC 3.4
24	11	2.8	499	2 A11472	heat-shock protein
25	11	2.8	500	2 AE1111	heat-shock protein
26	10	2.6	340	1 I64103	trypsin-like prote
27	10	2.6	362	2 T35287	probable secreted
28	10	2.6	382	2 H86930	probable secreted
29	10	2.6	389	2 B83089	Algw protein PA444

ALIGNMENTS

RESULT 1

C64647

serine proteinase (EC 3.4.21.-) - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jun-2000

C:Accession: C64647

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khatlani, H.G.; Glodek, A.; McKe

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: C64647

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-443 <TOM>

A:Cross-references: GB:AE000610; GB:AE000511; NID:g2314160; PIDN:AAD08063.1; PID:g231

C:Genetics:

A:Start codon: GTG

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

Query Match	100.0%;	Score 387;	DB 1;	Length 443;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 387;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 MIPKMERALGSGVIISKDGYIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSDDL 60

Db 57 MIPKMERALGSGVIISKDGYIVTNNHVIDGADKIKVTIPGSNKKEYSATLVGTDSDDL 116

Qy 61 AVIRITKDNLTIFKSDSDISVGLVFAIGNPFGVGSVTGGIVSALNKGSGINSYEN 120

Db 117 AVIRITKDNLTIFKSDSDISVGLVFAIGNPFGVGSVTGGIVSALNKGSGINSYEN 176

Qy 121 FIQTDASINPNSGALIDSRGLVGINTAIISKTGGNHGICFAIPSNMVDVTQLIKT 180

Db 177 FIQTDASINPNSGALIDSRGLVGINTAIISKTGGNHGICFAIPSNMVDVTQLIKT 236

Qy 181 GKIERGYLGVGLQDLSDQLQNSYDNKEGAVVISVEKDSAPKAGILVMDLITEVNGKKVK 240

Db 237 GKIERGYLGVGLQDLSDQLQNSYDNKEGAVVISVEKDSAPKAGILVMDLITEVNGKKVK 296

Qy 241 NTNELNLIGSMIPNQRVTLKVIKDKKERAFTLIAERKNPKKETTSAQNGAQLNGL 300

Db 297 NTNELNLIGSMIPNQRVTLKVIKDKKERAFTLIAERKNPKKETTSAQNGAQLNGL 356

Qy 301 QVEDLTQETKRSMLRSDDDVQGVLSOVNENSPAQAGFRQGNITTKIEVEVKSVADFNH 360

Db 357 QVEDLTQETKRSMLRSDDDVQGVLSOVNENSPAQAGFRQGNITTKIEVEVKSVADFNH 416

Qy 361 ALEKYKGPKRFLVLDLNOGYRIILVK 387

Db 417 ALEKYGKPRFLVDLNDQGYRIILVK 443
|||||

RESULT 2

H71936

Proteinase DO - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Sep-1999

C:Accession: H71936

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: H71936

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-476 <ARN>

A:Cross-references: GB:AE001474; GB:AE001439; NID:g4154929; PIDN:AA05980.1; PID:g415493

A:Experimental source: strain J99

C:Genetics:

A:Gene: htrA

C:Superfamily: Helicobacter serine proteinase

Query Match 24.8%; Score 96; DB 2; Length 476;

Best Local Similarity 100.0%; Pred. No. 8.8e-90;

Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 VTOLIKTKTERGLGVGLQDLSDQNSYDNKEGAVVISVEKDSPAKAGILVWDLITE 233

|||||

Db 263 VTOLIKTKTERGLGVGLQDLSDQNSYDNKEGAVVISVEKDSPAKAGILVWDLITE 322

|||||

QY 234 VNGKKVKNTELNLGSMPLPNORVTLKVIRDKKER 269

|||||

Db 323 VNGKKVKNTELNLGSMPLPNORVTLKVIRDKKER 358

RESULT 3

F81329

serine proteinase (proteinase DO) (EC 3.4.21.-) Cj1228c [imported] - Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C:Accession: F81329

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: F81329

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-472 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:g96968444; PIDN:CAB73482.1; PID:g9696866

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: htrA; Cj1228c

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

Query Match 5.2%; Score 20; DB 2; Length 472;

Best Local Similarity 100.0%; Pred. No. 6.4e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 YENFIOTDASINPGNSGGAL 137

|||||

Db 210 YENFIOTDASINPGNSGGAL 229

RESULT 4

DB2826

heat shock protein XF0285 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C:Accession: DB2826

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: DB2826

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-481 <SIM>

A:Cross-references: GB:AE003881; GB:AE003849; NID:g9105093; PIDN:AAF83098.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, L.E.A.; Carraro, D.M.; Carier

Briones, M.R.S.; Bueno, E.L.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Maracca, E.C.; Miyaki, C.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0285

C:Superfamily: Helicobacter serine proteinase

Query Match 4.7%; Score 18; DB 2; Length 481;

Best Local Similarity 100.0%; Pred. No. 7.3e-10;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 NFQTDSINPGNSGGAL 137

|||||

Db 222 NFQTDSINPGNSGGAL 239

RESULT 5

AI2811

serine proteinase htrA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 17-May-2002

C:Accession: AI2811

R:Wood, D.W.; Scubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: A82577; PMID:11743193

A:Accession: AI2811

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-468 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL42911.1; PID:gl7740366; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: htrA

A:Map position: circular chromosome

C:Superfamily: Helicobacter serine proteinase

Query Match 4.4%; Score 17; DB 2; Length 468;

Best Local Similarity 100.0%; Pred. No. 7.5e-09;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 FIQTDSINPGNSGGAL 137

|||||

Db 203 FIQTDSINPGNSGGAL 219

RESULT 6

B97590
htrA protein homolog [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 17-May-2002
C:Accession: B97590
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: B97590
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-495 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87675.1; PID:g15157032; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_3507
A:Map position: circular chromosome
C:Superfamily: Helicobacter serine proteinase

Query Match 4.4%; Score 17; DB 2; Length 495;
Best Local Similarity 100.0%; Pred. No. 7.9e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 FIOTDASINPGNSGGAL 137
|||||
Db 230 FIOTDASINPGNSGGAL 246

RESULT 7

H96956
serine protease Do (heat-shock protein) [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 16-Aug-2002
C:Accession: H96956
R:Noelling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; PMID:21359325; PMID:21359325
A:Accession: H96956
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78443.1; PID:g15023321; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0463
C:Superfamily: Escherichia coli trypsin-like proteinase; GLGF domain homology; trypsin h

Query Match 3.6%; Score 14; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 6.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 122 IOTDASINPGNSGG 135
|||||
Db 194 IOTDASINPGNSGG 207

RESULT 8

D89957
hypothetical protein SA1549 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 16-Aug-2002
C:Accession: D89957
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D89957

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-424 <KUR>

A:Cross-references: GB:BA000018; PID:g13701523; PIDN:BAW42817.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA1549

C:Superfamily: Escherichia coli trypsin-like proteinase; GLGF domain homology; trypsin

Query Match 3.1%; Score 12; DB 2; Length 424;

Best Local Similarity 100.0%; Pred. No. 0.00091;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 YIVTNNHVIDGA 33

|||||

Db 138 YIVTNNHVIDGA 149

RESULT 9

A64113
heat shock protein htrA - Haemophilus influenzae (strain Rd KW20)
N:Alternate names: periplasmic serine proteinase Do homolog
C:Species: Haemophilus influenzae
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: A64113
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman

; D.M.; Brandon, R.C.; White, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: A64113

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-466 <TIGR>

A:Cross-references: GB:U32805; GB:L42023; NID:g1574180; PIDN:AAK22906.1; PID:g1574189

C:Genetics:

A:Gene: htrA

C:Superfamily: Helicobacter serine proteinase

Query Match 3.1%; Score 12; DB 1; Length 466;

Best Local Similarity 100.0%; Pred. No. 0.00099;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 TNNHVIDGADKI 36

|||||

Db 117 TNNHVIDGADKI 128

RESULT 10

I40059
htrA-like protein - Brucella abortus
C:Species: Brucella abortus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: I40059
R:Patum, F.M.; Chevillat, N.F.; Morfitt, D.

Microb. Pathog. 17, 23-36, 1994

A:Title: Cloning, characterization and construction of htrA and htrA-like mutants of

A:Reference number: I40059; MUID:95165990; PMID:7861951

A:Accession: I40059

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-474 <EMBL>

A:Cross-references: EMBL-I40059; NID:g497154; PIDN:AAA70163.1; PID:g497155

C:Superfamily: Helicobacter serine proteinase

Query Match 3.1%; Score 12; DB 1; Length 474;

Best Local Similarity 100.0%; Pred. No. 0.001;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 INPGSGGALID 139
|||||
Db 215 INPGSGGALID 226

RESULT 11

proteinase DO (EC 3.4.21.-) precursor / heat shock protein htra - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Mar-2002
C:Accession: S45229; A64740; S01899; B35993
R:Fujita, N.
submitted to the EMBL Data Library, January 1994
A:Reference number: S45181
A:Accession: S45229
A:Molecule type: DNA
A:Residues: 1-474 <FULL>
A:CROSS-references: EMBL:D26562; NID:g473770; PIDN:BA05608.1; PID:g473819
A:Experimental source: strain K-12, substrain W3110
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, (1997)
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64740
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-474 <BLAT>
A:CROSS-references: GB:AE000125; GB:U00096; NID:g1786348; PIDN:AA073272.1; PID:g1786356;
A:Experimental source: strain K-12, substrain MGI655
R:Blipinska, B.; Sharma, S.; Georgopoulos, C.
Nucleic Acids Res. 16, 10053-10067, (1988)
A:Title: Sequence analysis and regulation of the htra gene of Escherichia coli: a sigma
A:Reference number: S01899; MUID:89057448; PMID:3057437
A:Accession: S01899
A:Molecule type: DNA
A:Residues: 1-9, 'R', '11-191, 'G', 193-466, 'RHLPNVAVISLNPFLKTKGRGSPYNL' <LIP>
A:CROSS-references: EMBL:X12457; NID:g41760; PIDN:CAA30997.1; PID:g41761
A:Experimental source: strain K-12
R:Wurgler, S.M.; Richardson, C.C.
Proc. Natl. Acad. Sci. U.S.A. 87, 2740-2744, (1990)
A:Title: Structure and regulation of the gene for dGTP triphosphohydrolase from Escherichia coli.
A:Reference number: A35993; MUID:90207273; PMID:2157212
A:Accession: B35993
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-9, 'R', '11-16 <WUR>
A:CROSS-references: GB:M31772; NID:g145733; PIDN:AAA23680.1; PID:g145736
C:Genetics:

A:Gene: htra
A:Map position: 4 min
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

1-26/Domain: signal sequence
127-474/Product: heat shock protein htra #status predicted <MAT>

Query Match 3.1%; Score 12; DB 2; Length 474;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
|||||
Db 263 GIGFAIPSNMVK 274

RESULT 12

proteinase DO (EC 3.4.21.-) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: A13349
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, N.; Mazur, M.; Goltsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis.
A:Reference number: AD3252; PMID:11756688
A:Accession: A13349
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <KUR>
A:CROSS-references: GB:AE008917; PIDN:AAL51964.1; PID:g17982723; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0783
A:Map position: I
A:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 3.1%; Score 12; DB 2; Length 474;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 INPGSGGALID 139
|||||
Db 215 INPGSGGALID 226

RESULT 13

proteinase DO (EC 3.4.21.-) precursor / heat shock protein htra - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C:Accession: E85500
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobeck, E.J.; Davis, N.W.; Ljm, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85500
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <STO>
A:CROSS-references: GB:AE005174; NID:g12512885; PIDN:AAG54465.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: htra
A:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 3.1%; Score 12; DB 2; Length 474;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSNMVK 171
|||||
Db 263 GIGFAIPSNMVK 274

RESULT 14

proteinase DO (EC 3.4.21.-) precursor / heat shock protein ECs0165 [imported] - Esche
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002
C:Accession: E90649
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90649
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <HAY>
A:CROSS-references: GB:BA000007; PIDN:BA033588.1; PID:g13359621; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509592
C:Genetics:

A:Gene: ECs0165
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 3.1%; Score 12; DB 2; Length 474;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 12; Conservative 0; Mismatches 0; Indels

QY 160 GIGFAIPSNMVK 171
|||
Db 263 GIGFAIPSNMVK 274

RESULT 15

S15337
heat shock protein htra - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S15337; S21327
R:Johnson, K.; Charles, I.; Dougan, G.; Pickard, D.; O'Gaora, P.; Costa, G.; Ali, T.; M
Mol. Microbiol. 5, 401-407, 1991
A:Title: The role of a stress-response protein in Salmonella typhimurium virulence.
A:Reference number: S15337; MUID:91251770; PMID:1645840
A:Accession: S15337
A:Molecule type: DNA
A:Residues: 1-475 <COS>
A:Cross-references: EMBL:X54548; NID:g47929; PID:g47930
C:Genetics:
A:Gene: htra
C:Superfamily: Helicobacter serine proteinase

Query Match 3.1%; Score 12; DB 1; Length 475;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 GIGFAIPSNMVK 171
 |||||
Db 264 GIGFAIPSNMVK 275

Search completed: November 18, 2002, 11:35:56
Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:30:11 ; Search time 40 Seconds
(without alignments)
1289.200 Million cell updates/sec

Title: US-09-895-913A-120
Perfect score: 387
Sequence: 1 MIPKERMERALSGVILSKD.....KPKRFLVLDLNGQYRIILVK 387

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	387	100.0	387	19 AAW98445	H. pylori GHPO 536
2	307	79.3	443	20 AAW98850	Antigen 1 from clu
3	288	74.4	288	20 AAW99984	Expressed antigen
4	165	42.6	451	20 AAW89977	Protein encoded by
5	105	27.1	187	20 AAW89817	Protein encoded by
6	13	3.4	474	22 AAG78605	Lawsonia intracell
7	12	3.1	475	12 AAR14625	S. typhimurium Htr
8	12	3.1	491	19 AAW56771	Escherichia coli s
9	11	2.8	168	22 ABG25263	Novel human diagno
10	11	2.8	380	23 ABP27963	Streptococcus poly

11	11	2.8	408	21	AA93768	The HtrA surface p
12	11	2.8	408	23	AB55490	Lactococcus lactis
13	11	2.8	409	21	AA91334	Group B Streptococ
14	11	2.8	409	23	ABP29849	Streptococcus poly
15	11	2.8	409	23	ABP30135	Streptococcus poly
16	11	2.8	455	22	AAG98946	E. coli growth and
17	11	2.8	500	23	AB448818	Listeria monocytog
18	11	2.8	503	17	AA77434	Heat shock protein
19	11	2.8	527	22	ABG25265	Novel human diagno
20	11	2.8	1084	22	ABG29110	Novel human diagno
21	10	2.6	76	23	ABP34815	Human protease-lik
22	10	2.6	238	22	ABG1538	Degp protease. St
23	10	2.6	245	20	AA935633	Chlamydia pneumonia
24	10	2.6	317	22	AAG82365	S. epidermidis ope
25	10	2.6	344	22	AAU60173	Propionibacterium
26	10	2.6	407	23	ABP27964	Streptococcus poly
27	10	2.6	413	23	ABP40266	Staphylococcus epi
28	10	2.6	433	21	AA939353	BAS011 protein se
29	10	2.6	433	21	AA939354	BAS011 protein se
30	10	2.6	433	21	AA939355	BAS011 protein se
31	10	2.6	464	20	AA904995	Mycobacterium spec
32	10	2.6	464	22	AA81131	Mycobacterium tube
33	10	2.6	488	23	AB90581	Chlamydia pneumonia
34	10	2.6	549	22	AAG81139	Mycobacterium tube
35	10	2.6	580	18	AA32431	Mycobacterium tube
36	10	2.6	580	18	AA32363	Mycobacterium tube
37	10	2.6	580	19	AAW81666	M. tuberculosis im
38	10	2.6	580	19	AAW64303	Mycobacterium tube
39	10	2.6	580	20	AA939105	M. tuberculosis an
40	10	2.6	580	20	AA938968	M. tuberculosis re
41	10	2.6	602	20	AA904996	Mycobacterium spec
42	9	2.3	234	22	AA861535	Degp protease. Es
43	9	2.3	236	22	AA861537	Degp protease. St
44	9	2.3	397	21	AA81772	Streptococcus pneu
45	8	2.1	10	22	AA861516	Degp protease cata

ALIGNMENTS

RESULT 1
AAW98445
ID AAW98445 standard: Protein; 387 AA.

AC AAW98445;

DT 31-MAR-1999 (first entry)

DE H. pylori GHPO 536 protein.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

KW peptic ulcer disease.

OS Helicobacter pylori.

PN WO9843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-US06371.

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

(HUMA-) HUMAN GENOME SCI INC.

(INNR) MERIEUX OIAVAX PASTEUR MERIEUX SERUMS.

PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX14164.


```
XX 27-APR-1998; 98WO-US08487.
XX
XX 14-OCT-1997; 97US-0061958.
XX 25-APR-1997; 97US-0045107.
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX Chow TP, Fry KE, Lim MY, McAtee CP;
XX WPI; 1999-009433/01.
XX
XX New Helicobacter pylori antigens and related nucleic acid sequences
XX - useful in serological diagnosis and protective vaccines, providing
XX long-lasting immune response
XX
XX Claim 16; Page 322-323; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
XX that is characterised by immunoreactivity with H. pylori-positive
XX antisera. The proteins are highly immunogenic and induce a long-lasting
XX immune response that persists even after antimicrobial treatment. In
XX antibody-detection assays, on sera, plasma, urine, saliva etc., they are
XX highly sensitive and specific. The specification also describes 69
XX previously unrecognised immunogenic cluster families. H. pylori antigens
XX are used to detect H. pylori-specific antibodies, for diagnosing
XX infection or to confirm eradication of infection, and in vaccines to
XX protect against H. pylori infection and related diseases (gastritis,
XX peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX Sequence 288 AA;
XX
XX Query Match 74.4%; Score 288; DB 20; Length 288;
XX Best Local Similarity 100.0%; Pred. No. 6.1e-288;
XX Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 100 VTQGVSAALNGKIGINSYENFIQDASINPGNSGGALIDSRGLVGINTAISKGTGNH 159
XX Db 1 VTQGVSAALNGKIGINSYENFIQDASINPGNSGGALIDSRGLVGINTAISKGTGNH 60
XX
XX QY 160 GIGFAIPSNMVKDVTQLIKTKIERGVGLVGLQDLSGLQNSYDNKGAIVISYEKDSK 219
XX Db 61 GIGFAIPSNMVKDVTQLIKTKIERGVGLVGLQDLSGLQNSYDNKGAIVISYEKDSK 120
XX
XX QY 220 AKKAGILWDLITEYNGKKVNTNELNLIGSMLPNQVTLKVIKDKKERAFTLTLAERK 279
XX Db 121 AKKAGILWDLITEYNGKKVNTNELNLIGSMLPNQVTLKVIKDKKERAFTLTLAERK 180
XX
XX QY 280 NPNKKTISAQNGAQQQLNGLQVEDLTQETKRSMLSDVQGVLYSQVNENSPAQAGFR 339
XX Db 181 NPNKKTISAQNGAQQQLNGLQVEDLTQETKRSMLSDVQGVLYSQVNENSPAQAGFR 240
XX
XX QY 340 QGNITTKIEEVKSVADFNHAEKYGKPKRFLVDLNOGYRIILVK 387
XX Db 241 QGNITTKIEEVKSVADFNHAEKYGKPKRFLVDLNOGYRIILVK 288
XX
XX RESULT 4
XX AAW89977
XX ID AAW89977 standard; Protein; 451 AA.
XX
XX AC AAW89977;
XX
XX DT 18-FEB-1999 (first entry)
XX
XX DE Protein encoded by clone b8.
XX
XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
XX OS Helicobacter pylori.
XX
XX PN W09849314-A2.
```

```
XX 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08487.
XX
XX 14-OCT-1997; 97US-0061958.
XX 25-APR-1997; 97US-0045107.
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX Chow TP, Fry KE, Lim MY, McAtee CP;
XX WPI; 1999-009433/01.
XX N-PSDB; AAV90921.
XX
XX New Helicobacter pylori antigens and related nucleic acid sequences
XX - useful in serological diagnosis and protective vaccines, providing
XX long-lasting immune response
XX
XX Claim 15; Page 317-318; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
XX that is characterised by immunoreactivity with H. pylori-positive
XX antisera. The proteins are highly immunogenic and induce a long-lasting
XX immune response that persists even after antimicrobial treatment. In
XX antibody-detection assays, on sera, plasma, urine, saliva etc., they are
XX highly sensitive and specific. The specification also describes 69
XX previously unrecognised immunogenic cluster families. H. pylori antigens
XX are used to detect H. pylori-specific antibodies, for diagnosing
XX infection or to confirm eradication of infection, and in vaccines to
XX protect against H. pylori infection and related diseases (gastritis,
XX peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX Sequence 451 AA;
XX
XX Query Match 42.6%; Score 165; DB 20; Length 451;
XX Best Local Similarity 100.0%; Pred. No. 5.2e-161;
XX Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MIPKMERALGSGVLIISKDGYIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTDSDDL 60
XX Db 65 MIPKMERALGSGVLIISKDGYIVTNNHVIDGADKIKVTIPGSNKEYSATLVGTDSDDL 124
XX
XX QY 61 AVIRITTKNLPTIKFSDSNDISVGLDVFAGNPFVGESVTOGIVSALNKGIGINSYEN 120
XX Db 125 AVIRITTKNLPTIKFSDSNDISVGLDVFAGNPFVGESVTOGIVSALNKGIGINSYEN 184
XX
XX QY 121 FIQTASINPGNSGGALIDSRGLVGINTAISKGTGNHGIGFAI 165
XX Db 185 FIQTASINPGNSGGALIDSRGLVGINTAISKGTGNHGIGFAI 229
XX
XX RESULT 5
XX AAW89817
XX ID AAW89817 standard; Protein; 187 AA.
XX
XX AC AAW89817;
XX
XX DT 18-FEB-1999 (first entry)
XX
XX DE Protein encoded by clone C7.
XX
XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
XX OS Helicobacter pylori.
XX
XX PN W09849314-A2.
XX
XX PD 05-NOV-1998.
XX
XX PF 27-APR-1998; 98WO-US08487.
```

PR 14-OCT-1997; 97US-0061958.
PR 23-APR-1997; 97US-0045107.
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX Chow TP, Fry KE, Lim MY, McAtee CP;
XX WPI; 1999-009433/01.
XX N-PSDB; AAV90552.
XX
XX New helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT - long-lasting immune response
XX
XX Claim 15; Page 103; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
XX that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX Sequence 187 AA;
XX
XX Query Match 27.1%; Score 105; DB 20; Length 187;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-99;
XX Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 203 YDNKEGAVVLSVEKDSAPKAGILVLDLITEVNGKKVKNTELRNLIGSMLPNQVTLKV 262
DB 3 YDNKEGAVVLSVEKDSAPKAGILVLDLITEVNGKKVKNTELRNLIGSMLPNQVTLKV 62
XX
QY 263 IRDKKERAFTLTLAERKNPNKKTETISAQNGAQGLQGLVEDLTQ 307
DB 63 IRDKKERAFTLTLAERKNPNKKTETISAQNGAQGLQGLVEDLTQ 107
XX
XX
XX RESULT 6
XX AG78605
XX ID AAG78605 standard; Protein; 474 AA.
XX
XX AC AAG78605;
XX
XX 20-NOV-2001 (first entry)
XX
XX Lawsonia intracellularis protein SEQ ID NO: 7.
XX
XX HtrA; PonA; HypC; Yefw; ABC1; Omp100; Lawsonia intracellularis infection;
XX vaccine.
XX
XX Lawsonia intracellularis.
XX
XX JP2001169787-A.
XX
XX 26-JUN-2001.
XX
XX 20-OCT-2000; 2000JP-0320736.
XX
XX 22-OCT-1999; 99US-0160922.
XX (PFI2) PFIZER PROD INC.
XX
XX WPI; 2001-592540/67.
XX
XX Lawsonia intracellularis polynucleotide and encoded protein, used to
PT prevent Lawsonia intracellularis infection -
XX

PS Claim 12; Page 47-49; 67pp; Japanese.
XX
XX The present invention provides isolated polynucleotides encoding HtrA,
CC PonA, HypC, LysS, Yefw, ABC1 or Omp100 protein of Lawsonia
CC intracellularis. The sequences can be used in vaccines for the prevention
CC of Lawsonia intracellularis infection. The present sequence is a protein
CC of the invention.
XX
XX SQ Sequence 474 AA;
XX
XX Query Match 3.4%; Score 13; DB 22; Length 474;
XX Best Local Similarity 100.0%; Pred. No. 0.00022;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 123 QTDASINPGNSGG 135
DB 207 QTDASINPGNSGG 219
XX
XX RESULT 7
XX AAR14625
XX ID AAR14625 standard; Protein; 475 AA.
XX
XX AC AAR14625;
XX
XX 14-JAN-1992 (first entry)
XX
XX S. typhimurium HtrA protein.
XX
XX TnpHoA mutagenesis; signal peptide; stress protein;
XX heat shock protein; degp.
XX
XX Salmonella typhimurium strain C5.
XX
XX WO9115572-A.
XX
XX 17-OCT-1991.
XX
XX 23-MAR-1991; 91WO-GB00484.
XX
XX 30-MAR-1990; 90GB-0007194.
XX
XX (WELL) WELLCOME FOUNDATION LTD.
XX
XX Dougan G, Charles IG, Hormaeche CE, Johnson KS, Chatfield SN;
XX WPI; 1991-325215/44.
XX N-PSDB; AAO14416.
XX
XX Attenuated microorganism useful in live vaccines - attenuated by
XX mutation in DNA sequence encoding e.g. a heat shock protein
XX
XX Disclosure; Fig 1; 28pp; English.
XX
XX TnpHoA mutagenesis was used in the mouse virulent S. typhimurium
XX strain C5. Mutants were selected likely to harbour lesions in genes
XX that have a signal peptide sequence, i.e. proteins likely to be
XX targeted through a bacterial membrane. Isolation of the DNA
XX flanking the TnpHoA insertion identified the gene (htrA) that has
XX been insertionally activated. Comparison of the translated protein
XX sequence showed that it shared 88% homology with the sequence of the
XX htrA product from E. coli.
XX
XX SQ Sequence 475 AA;
XX
XX Query Match 3.1%; Score 12; DB 12; Length 475;
XX Best Local Similarity 100.0%; Pred. No. 0.0024;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 160 GIGFAIPSNMVK 171
DB 264 GIGFAIPSNMVK 275

RESULT 8
AAW56771
ID AAW56771 standard; Protein; 491 AA.
XX
AC AAW56771;
XX
DT 13-OCT-1998 (first entry)
XX
DE Escherichia coli serine protease htrA.
XX
XX PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease; htrA;
KW serine protease; neurodegeneration; predisposition; diagnosis.
XX
OS Escherichia coli.
XX
PN EP828003-A2.
XX
PD 11-MAR-1998.
XX
XX 26-AUG-1997; 97EP-0306501.
XX
PR 13-DEC-1996; 96US-0032875.
PR 06-SEP-1996; 96US-0025436.
PR 25-OCT-1996; 96US-0027873.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Browne MJ, Clinkenbeard HE, Creasy CL, Karran EH;
PI Livl GP, Southan CB;
XX
DR WPI; 1998-1611101/15.
DR N-PSDB; AAV29528.
XX
XX Nucleic acids encoding human serum protease protein(s) - used for
PT diagnosing pre-disposition to Alzheimer's disease, etc.
XX
PS Example 2; Page 28-29; 65pp; English.
XX
XX The sequence is that of the serine protease htrA which was used
CC in the isolation of PSPI.
XX
SQ Sequence 491 AA;
Query Match 3.1%; Score 12; DB 19; Length 491;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 160 GIGFAIPSNMVK 171
Db 263 GIGFAIPSNMVK 274
|||||
RESULT 9
ABG25263
ID ABG25263 standard; Protein; 168 AA.
XX
AC ABG25263;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #25254.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS89450.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 55622; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 168 AA;
Query Match 2.8%; Score 11; DB 22; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 119 ENFIQTDSIN 129
Db 100 ENFIQTDSIN 110
|||||
RESULT 10
ABP27963
ID ABP27963 standard; Protein; 380 AA.
XX
AC ABP27963;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 5102.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 XX WPI: 2002-352536/38.
 DR N-PSDB; ABN68594.
 DR
 XX
 XX New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 XX for detecting a compound that binds to the protein -
 XX
 XX Claim 1; Page 3673; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 XX Sequence 380 AA;
 SQ
 Query Match 2.8%; Score 11; DB 23; Length 380;
 Best Local Similarity 100.0%; Pred. No. 0.021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 INPGNSGGALI 138
 DB 203 INPGNSGGALI 213
 |||||
 RESULT 11
 AAY93768
 ID AAY93768 standard; Protein; 408 AA.
 XX
 AC AAY93768;
 XX
 DT 03-OCT-2000 (first entry)
 XX
 XX The HtrA surface protease of Lactococcus lactis.
 XX
 DE Catalytic site: protease; HtrA: surface protease; fermented product;
 KW vaccine; dietary protein; cheese production; digestive enzyme.
 KW
 XX Lactococcus lactis.
 OS
 XX Key Location/Qualifiers
 FH Domain 10..26
 FT /note= "transmembrane domain"
 FT Disulfide-bond 119..129
 FT /note= "catalytic domain"
 FT Domain 157..164
 FT /note= "catalytic domain"
 FT Domain 237..254
 FT /note= "catalytic domain"
 FT

XX WO200039309-A1.
 PN
 XX
 PD 06-JUL-2000.
 XX
 XX 23-DEC-1999; 99WO-FR03270.
 PF
 XX
 PR 24-DEC-1998; 98FR-0016462.
 PR
 XX
 PA (INRG) INST NAT RECH AGRONOMIQUE.
 XX
 XX Poquet I, Gruss A, Bolotine A, Sorokine A;
 PI WPI: 2000-465747/40.
 DR N-PSDB; AAA471161.
 DR
 XX
 XX Bacteria deficient in activity of protease HtrA, useful for production
 PT of proteins, e.g. vaccinating antigens, that are exported from the
 PT cell, provide increased yield -
 PT
 XX Disclosure; Fig 1; 42pp; French.
 PS
 XX The present sequence represents a HtrA surface antigen from Lactococcus
 CC lactis subsp. lactis. The specification describes a method for the
 CC production of a protein. The method comprises culturing a bacterial
 CC strain that expresses the protein and is prepared from a Gram-positive
 CC species with a genome no larger than 3.2 Mb by mutational inactivation
 CC of the HtrA surface protease, and recovering the protein exported from
 CC the cells. Inactivation of HtrA almost completely eliminates degradation
 CC of exported proteins, so increases yields of the protein and prevents
 CC contamination by proteolytic degradation products (facilitating
 CC purification at reduced cost). The bacterial strain is used for
 CC preparation of fermented products, therapeutic proteins (especially
 CC vaccines) or dietary proteins (e.g. for the production of cheeses or
 CC enzymes for facilitating digestion).
 XX
 XX Sequence 408 AA;
 SQ
 Query Match 2.8%; Score 11; DB 21; Length 408;
 Best Local Similarity 100.0%; Pred. No. 0.022;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 INPGNSGGALI 138
 DB 234 INPGNSGGALI 244
 |||||
 RESULT 12
 ABB55490
 ID ABB55490 standard; Protein; 408 AA.
 XX
 AC ABB55490;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 XX Lactococcus lactis protein htrA.
 DE
 DE Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 KW
 XX Lactococcus lactis I11403.
 OS
 XX FR2807446-A1.
 PN
 XX
 PD 12-OCT-2001.
 XX
 XX 11-APR-2000; 2000FR-0004630.
 PF
 XX
 PR 11-APR-2000; 2000FR-0004630.
 PR
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
 PA
 XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 PI

DR WPI; 2002-043418/06.
XX New nucleotide sequence useful in the identification or Lactococcus
PT lactis and related species -
XX
XX
PS Claim 6; SEQ ID No 2192; 2504pp; French.
XX
XX The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 408 AA;
Query Match 2.8%; Score 11; DB 23; Length 408;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 128 INPGNSGGALI 138
|||||
Db 234 INPGNSGGALI 244
|||||
RESULT 13
AAAY91334
ID AAY91334 standard; Protein; 409 AA.
XX
AC AAY91334;
XX
XX 30-MAY-2000 (first entry)
XX
XX Group B Streptococcus protein sequence SEQ ID NO:67.
XX
XX Group B Streptococcus; Streptococcus agalactiae; protein antigen;
KW vaccine; screening; immunogen; detection; diagnosis; infection;
KW antibody; antibody; antibacterial.
XX
OS Streptococcus agalactiae.
XX
PN WO200006736-A2.
XX
PD 10-FEB-2000.
XX
XX 27-JUL-1999; 99WO-GB02444.
XX
XX 27-JUL-1998; 98GB-0016335.
XX 19-MAR-1999; 99US-0125163.
XX (MICR-) MICROBIAL TECHNIQS LTD.
XX
XX Le Page RWF, Wells JM, Hanniffy SB;
XX
XX WPI; 2000-195299/17.
XX
XX New Group B Streptococcus protein, useful as vaccine, for diagnosis of
PT Streptococcal infections and for screening of antibodies or antibodies
XX
XX Claim 1; Fig 1; 123pp; English.
XX
XX AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given
CC in AAY91275 to AAY91343) isolated from group B Streptococcus (GBS), also
CC known as Streptococcus agalactiae. The GBS polynucleotides and
CC polypeptides have antibacterial activity. Immunogenic compositions
CC comprising GBS polynucleotides or polypeptides can be used as vaccines
CC and for the treatment or prophylaxis of GBS infection. The
CC polynucleotides and polypeptides can also be used in the detection of GBS

CC and for screening DNA encoding bacterial cell envelope associated or
CC secreted antigens in gram positive bacteria. AAA05873 to AAA05941
CC represent primers used in the exemplification of the present invention.
XX
XX
SQ Sequence 409 AA;
Query Match 2.8%; Score 11; DB 21; Length 409;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 128 INPGNSGGALI 138
|||||
Db 232 INPGNSGGALI 242
|||||
RESULT 14
ABP29849
ID ABP29849 standard; Protein; 409 AA.
XX
AC ABP29849;
XX
XX 02-JUL-2002 (first entry)
XX
XX Streptococcus polypeptide SEQ ID NO 8874.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000QB-0026333.
PR 24-NOV-2000; 2000QB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masigiani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR N-PSDB; ABN70480.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX Claim 1; Page 4005; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

```

Query Match          2.8%; Score 11; DB 23; Length 409;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 INPGNSGGALI 138
      |||||
Db 232 INPGNSGGALI 242

```

XX	Sequence	409 AA;
SQ		
	Query Match	2.8%; Score 11; DB 23; Length 409;
	Best Local Similarity	100.0%; Pred. No. 0.022;
	Matches 11; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	128 INFGNSGGALI 138	
Db	232 INFGNSGGALI 242	
	ABP30135;	
XX	02-JUL-2002 (first entry)	
XX		
DT		
DE	Streptococcus polypeptide SEQ ID NO 9446.	
XX		
KW	Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;	
KW	group A streptococcus; Streptococcus pyogenes; antibacterial;	
KW	antiinflammatory; infection; vaccine; meningitis; gene therapy.	
XX		
XX	Streptococcus agalactiae.	
OS		
XX	WO200234771-A2.	
PN	02-MAY-2002.	
PD		
XX	29-OCT-2001; 2001WO-GB04789.	
PF		
XX	27-OCT-2000; 2000GB-0026333.	
PR	24-NOV-2000; 2000GB-0028727.	
PR	07-MAR-2001; 2001GB-0005640.	
XX		
XX	(CHIR-) CHIRON SPA.	
PA	(GENO-) INST GENOMIC RES.	
XX	Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;	
PI	Tettelin H;	
DR	WPI; 2002-352536/38.	
N-PSDB; ABN70766.		
XX	New Streptococcus protein for the treatment or prevention of infection	
PT	or disease caused by Streptococcus bacteria, such as meningitis, and	
PT	for detecting a compound that binds to the protein -	
XX	Claim 1; Page 4070; 4525pp; English.	
XX	The invention relates to a protein (ABP25413-ABP30895) from group B	
CC	streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS	
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in	
CC	the specification. The proteins have antibacterial and antiinflammatory	
CC	activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and	
CC	antibodies that bind (I) are used in the manufacture of medicaments for	
CC	the treatment or prevention of infection or disease caused by	
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.	
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a	
CC	biological sample. (I) is used to determine whether a compound binds to	
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be	
CC	used as a vaccine or diagnostic composition. The disease caused by	
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic	
CC	acid encoding (I) may be used to recombinantly produce (I) and may be	
CC	used in gene therapy. Antibodies to (I) are used for affinity	
CC	chromatography, immunoassays, and distinguishing/identifying	
CC	Streptococcus proteins.	
XX		
SQ	Sequence 409 AA;	

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:33:16 ; Search time 13 seconds
(without alignments)
1234.718 Million cell updates/sec

Title: US-09-895-913A-120

Perfect score: 387

Sequence: 1 MKPKRMERALGSGVIISKD.....KPKRFLVLDLNOGYRIILVK 387

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	12	3.1	456	HTOA_HAEIN	P45129 haemophilus
2	12	3.1	474	DEGP_ECOLI	P09376 escherichia
3	12	3.1	475	DEGP_SALTY	P29882 salmonella
4	11	2.8	408	HTRA_LACLA	Q91a06 lactococcus
5	11	2.8	455	DEGP_ECOLI	P39099 escherichia
6	11	2.8	503	DEGP_BARHE	P54925 bartonella
7	10	2.6	340	DEGS_HAEIN	P44947 haemophilus
8	10	2.6	400	YXA_BACSU	P39688 bacillus su
9	10	2.6	413	HTRA_LACHE	Q924h7 lactobacilli
10	10	2.6	488	DEGP_CHLPN	Q926t0 chlamydia p
11	9	2.3	437	DEGL_ARATH	Q22609 arabidopsis
12	9	2.3	449	HTRA_BACSU	Q34358 bacillus su
13	9	2.3	478	DEGP_BUCAL	P57322 buchnera ap
14	9	2.3	478	DEGP_BUCAL	O85291 buchnera ap
15	9	2.3	508	DEGP_RICCN	Q924a1 rickettsia
16	9	2.3	513	DEGP_RICPR	O05942 rickettsia
17	8	2.1	96	RL23_BACHD	Q92912 bacillus ha
18	8	2.1	256	VBRL_PHV	O06927 pepper huas
19	8	2.1	256	VBRL_TGVV	P03564 tomato gold
20	8	2.1	239	RL5_BOMMO	O76190 bombyx mori
21	8	2.1	338	G3P_SCLSC	Q96u88 sclerotinia
22	8	2.1	448	DEGP_ARATH	Q91u10 arabidopsis
23	8	2.1	458	YVTA_BACSU	Q91911 bacillus su
24	8	2.1	497	DEGP_CHLMU	Q9197 chlamydia m
25	8	2.1	497	DEGP_CHLTR	P18584 chlamydia t
26	7	1.8	54	ATP8_BRAFL	O47427 branchiost
27	7	1.8	54	ATP8_BRALA	O21003 branchiost
28	7	1.8	67	RL29_BACHD	Q929k6 bacillus ha
29	7	1.8	113	TSQA_YEAST	O69232 buchnera ap
30	7	1.8	195	POA_BUCAP	O04120 saccharomyc
31	7	1.8	221	PP_LAMBD	P03772 bacterioph
32	7	1.8	256	VBRL_PYMVV	P27268 potato yell
33	7	1.8	269	SAPT_HAEIN	P45289 haemophilus

RESULT 1
HTOA_HAEIN STANDARD; PRT; 466 AA.
AC P45129;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable periplasmic serine protease do/hhoA-like precursor
DE (EC 3.4.21.-)
GN H11259
OS Haemophilus influenzae
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -|- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -|- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA
(PROTEASE DO) AND HHOA.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -|- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32805; AAC22906.1; -;
CC MEROPS: S01.274; -;
CC TIGR: H11259; -;
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001940; Protease2C.
CC InterPro: IPR001254; Ser.protease_Try.
CC Pfam: PF00089; trypsin; 1.
CC Pfam: PF00595; PDZ; 2.
CC PRINTS: PR00834; PROTEASES2C.
CC SMART: SM00228; PDZ; 2.
CC PROSITE: PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Periplasmic; Repeat; Signal;

ALIGNMENTS

DR EMBL; AP002550; BAB33588.1; -.
 DR EMBL; M29955; AAA23717.1; -.
 DR EMBL; M31772; AAA23680.1; -.
 DR PIR; S01899; S01899.
 DR PIR; B35993; B35993.
 DR MEROPS; S01.273; -.
 DR SWISS-2DPAGE; P09376; COLI.
 DR EcoGene; EG10463; degp.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 2.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 DR Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 474
 FT DOMAIN 280 371
 FT DOMAIN 377 466
 FT ACT_SITE 131 131
 FT ACT_SITE 161 161
 FT ACT_SITE 236 236
 FT CONFLICT 10 10
 FT CONFLICT 46 46
 FT CONFLICT 192 192
 FT CONFLICT 467 474
 FT SEQUENCE 474 AA; 49354 MW; 54825967486D5F CRC64;
 Query Match 3.1%; Score 12; DB 1; Length 474;
 Best Local Similarity 100.0%; Pred. No. 0.00024;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 160 GIGFAIPSNMVK 171
 DB 263 GIGFAIPSNMVK 274
 |||||
 RESULT 3
 DEGP_SALTY STANDARD; PRT; 475 AA.
 AC P26982;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protease do precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR PTD OR STM0209.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OR NCBI_taxid=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C5;
 RX MEDLINE=91251770; PubMed=1645840;
 RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
 RA Ali T., Miller I., Hornaechne C.;
 RT "The role of a stress-response protein in Salmonella typhimurium
 RT virulence."
 RL Mol. Microbiol. 5:401-407(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2.";
 RL Nature 413:852-856(2001).
 CC FUNCTION:-SERINE-PROTEASE-THAT-IS-REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
 CC SPECIFICITY WITH HHOA/DEGO.
 CC -!- INDUCTION: BY HEAT SHOCK.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X54548; CAA38420.1; -.
 DR EMBL; AE008704; AAL19173.1; -.
 DR PIR; S15337; S15337.
 DR PIR; S21327; S21327.
 DR MEROPS; S01.273; -.
 DR StyGene; SG10173; degp.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 2.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 DR Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 475
 FT DOMAIN 281 372
 FT DOMAIN 378 467
 FT ACT_SITE 132 132
 FT ACT_SITE 162 162
 FT ACT_SITE 237 237
 FT SEQUENCE 475 AA; 49315 MW; 8656858F3C1A289F CRC64;
 Query Match 3.1%; Score 12; DB 1; Length 475;
 Best Local Similarity 100.0%; Pred. No. 0.00024;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 160 GIGFAIPSNMVK 171
 DB 264 GIGFAIPSNMVK 275
 |||||
 RESULT 4
 HTRA_LACIA STANDARD; PRT; 408 AA.
 ID Q9LA06;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Serine protease dc-like htra (EC 3.4.21.-) (HtraL1).
 GN HTRA OR L12136.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OR NCBI_taxid=1360;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC STRAIN=ILL403;
 RX MEDLINE=20177820; PubMed=10712686;
 RA Poquet I., Saint V., Seznec E., Smoes N., Bolotin A., Gruss A.;
 RT "Htra is the unique surface housekeeping protease in Lactococcus
 RT lactis and is required for natural protein processing."
 RL Mol. Microbiol. 35:1042-1051(2000).
 RN [2]

GN DEGP OR HTRA.
 OS Bacteraia; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bartonellaceae; Bartonella.
 OX NCBI_TaxID=38323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Houston-1;
 RX MEDLINE=94299828; PubMed=8027347;
 RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
 RT Goral S., Hager C., Edwards K.;
 RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch
 disease patients by PCR."
 RL J. Clin. Microbiol. 32:942-948 (1994).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L20127; AAA97430.1; -
 DR MEROPS; S01.273; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 2.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 DR Hydroxylase; Serine protease; Periplasmic; Repeat; Signal.
 KW SIGNAL 1 18
 FT CHAIN 19 503
 FT LIKE 19 503
 FT DOMAIN 286 357
 FT PDZ 1.
 FT PDZ 2.
 FT ACT_SITE 143 143
 FT CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 173 173
 FT CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 247 247
 FT CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 503 AA; 54114 MW; 6CD9F4743282AF9E CRC64;
 Query Match 2.8%; Score 11; DB 1; Length 503;
 Best Local Similarity 100.0%; Pred. No. 0.0028;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 20 DGYIVTNHVI 30
 Db 135 DGYIVTNHVI 145
 |||||
 RESULT 7
 DEGS_HAEIN
 ID DEGS_HAEIN STANDARD; PRT; 340 AA.
 AC P44947;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protease degs precursor (EC 3.4.21.-).
 GN DEGS OR HHOB OR HTRH OR H10945.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Hd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd."
 RL Science 269:496-512 (1995).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32775; AAC2:599.1; -
 DR MEROPS; S01.275; -
 DR TIGR; H10945; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR Hydroxylase; Serine protease; Periplasmic; Signal; Complete proteome.
 KW SIGNAL 1 22
 FT CHAIN 23 340
 FT DOMAIN 251 323
 FT PDZ.
 FT ACT_SITE 92 92
 FT CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 122 122
 FT CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 340 AA; 36039 MW; 69EA452DF5A10649 CRC64;
 Query Match 2.6%; Score 10; DB 1; Length 340;
 Best Local Similarity 100.0%; Pred. No. 0.022;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 120 NFQIQTASIN 129
 Db 184 NFQIQTASIN 193
 |||||
 RESULT 8
 YXXA_BACSU
 ID YXXA_BACSU STANDARD; PRT; 400 AA.
 AC P39668;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical serine protease yxxa (EC 3.4.21.-).
 GN YXXA OR YVCK.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kasahara Y., Nakai S., Yoshikawa H., Ogasawara N.;
 RT "38kb sequence between gntZ and trnY of B. subtilis genome."
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

1995

```
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Mosser J., Albertini A.M., Alloni G.,
RA Azevedo V., Berto L.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriass R., Bertsier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entlan K.D., Errington J., Fabret C., Ferrali E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viarl A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RN SEQUENCE OF 317-400 FROM N.A.
RC STRAIN=168; Marburg;
RX MEDLINE=94156824; PubMed=8113162;
RA Calogero S., Gardan R., Glaser P., Schweitzer J., Rapoport G.,
RA Debarbouille M.,
RT "RocR, a novel regulatory protein controlling arginine utilization in
RT Bacillus subtilis, belongs to the NtrC/NifA family of transcriptional
RT activators."
RL J. Bacteriol. 176:1234-1241(1994).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D78193; BAAL1295.1; -.
DR EMBL: Z99124; CAB16073.1; -.
DR EMBL: L22006; -; NOT_ANNOTATED_CDS.
DR MEROPS: S01.273; -.
DR Subtilist; BG11054; yYxA.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50106; PDZ; FALSE_NEG.
KW Hypothetical protein; Hydrolase; Serine protease; Transmembrane;
KW Complete proteome.
FT ACT_SITE 129 129 POTENTIAL.
FT TRANSMEM 22 42
FT ACT_SITE 159 159 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 159 159 CHARGE RELAY SYSTEM (POTENTIAL).
FT -----
FT ACT_SITE 243 243 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 400 AA; 42788 MW; 91A6E0E897B6F64C CRC64;
Query Match
Best Local Similarity 2.6%; Score 10; DB 1; Length 400;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 128 INPGNSGGAL 137
Db 238 INPGNSGGAL 247
|||||
RESULT 9
HTRA_LACHE
ID HTRA_LACHE STANDARD; PRT; 413 AA.
AC Q9Z4H7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (BC 3.4.21.-).
GN HTRA.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=537;
RX MEDLINE=98047559; PubMed=9829922;
RA Smeds A., Varmanen P.K., Palva A.M.;
RT "Molecular characterization of a stress-inducible gene from
RT Lactobacillus helveticus."
RL J. Bacteriol. 180:6143-6153(1998).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ005672; CAA06668.1; -.
DR MEROPS: S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolase; Serine protease; Transmembrane.
FT TRANSMEM 20 40
FT DOMAIN 141 302 POTENTIAL.
FT DOMAIN 305 401 CATALYTIC.
FT ACT_SITE 150 150 PDZ.
FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 413 AA; 42647 MW; B16B677991C88707 CRC64;
Query Match
Best Local Similarity 2.6%; Score 10; DB 1; Length 413;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 GYIVTNHHVI 30
Db 143 GYIVTNHHVI 152
|||||
RESULT 10
DEGP_CHLPN
```

DEGP_CHLPN STANDARD; PRT; 488 AA.
Q9Z6T0; Q9JQD7; Q9K1W4;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Probable serine protease do-like precursor (EC 3.4.21.-).
DEGP OR HTRA OR CPN0979 OR CP0877.
Chlamydia pneumoniae (Chlamydia pneumoniae).
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83558;
[1]
SEQUENCE FROM N.A.
STRAIN=CWL029;
MEDLINE=99206606; PubMed=10192388;
Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
Nat. Genet. 21:385-389(1999).
[2]
SEQUENCE FROM N.A.
STRAIN=AR39;
MEDLINE=20150255; PubMed=10684935;
Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
Ginn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
Eisen J., Fraser C.M.;
"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39."; Nucleic Acids Res. 28:1397-1406(2000).
[3]
SEQUENCE FROM N.A.
STRAIN=J138;
MEDLINE=20330349; PubMed=10871362;
Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
"Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA."; Nucleic Acids Res. 28:2311-2314(2000).
[1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
[1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; AE001678; AAD19116.1; -
EMBL; AE002246; AAF38665.1; -
EMBL; AP002548; BAA99186.1; -
MEROPS; S01.273; -
PHCI-2DPAGE; Q9Z6T0; -
TIGR; CP0877; -
InterPro; IPR001478; PDZ.
InterPro; IPR001940; Protease2C.
InterPro; IPR001254; Ser_protease_Try.
Pfam; PF00089; trypsin; 1.
Pfam; PF00595; PDZ; 2.
SMART; SM00228; PDZ; 2.
SMART; SM00228; PDZ; 2.
PROSITE; PS01016; PDZ; 2.
Hydrolase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 488 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 119 280 CATALYTIC.
FT DOMAIN 281 372 PDZ 1.
FT DOMAIN 388 476 PDZ 2.
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 164 164 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).

SQ SEQUENCE 488 AA; 52311 MW; 0EE7E0F88F106F49 CRC64;
Query Match 2.6%; Score 10; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 DGYIVTNNHV 29
Db 126 DGYIVTNNHV 135

RESULT 11
DEGI_ARATH
ID DEGI_ARATH STANDARD; PRT; 437 AA.
AC 022609; O9L85;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).
GN DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=98175982; PubMed=9507020; Cook M., Adam Z.;
Itzhaki H., Naveh I., Lindahl M., Lindahl M., Lindahl M., Lindahl M.,
"Identification and characterization of Degp, a serine protease
associated with the luminal side of the thylakoid membrane."; J.
J. Biol. Chem. 273:7094-7098(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
"Structural analysis of Arabidopsis thaliana chromosome 3. II.
Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
and BAC clones."; DNA Res. 7:217-221(2000).
RN [3]
RP SEQUENCE OF 104-111.
RC STRAIN=cv. Columbia;
RA Kieselbach T., Bystedt M., Schroeder W.P.;
Submitted (JUL-2000) to the SWISS-PROT data bank.
CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,
CC CAN DEGRADE BETA-CASEIN.
CC -1- ENZYME REGULATION: INHIBITED BY PHENYL METHYL SULFONYL FLUORIDE AND
CC O-PHENANTHROLINE.
CC -1- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID
CC MEMBRANE.
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; AF028842; AAC39436.1; -
EMBL; AP000371; BAB02539.1; -
EMBL; AP001302; BAB02539.1; JOINED.
MEROPS; S01.279; -
InterPro; IPR001478; PDZ.
InterPro; IPR001940; Protease2C.
InterPro; IPR001254; Ser_protease_Try.
Pfam; PF00089; trypsin; 1.

DR Pfam: PF00595; PDZ; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS0106; PDZ; 1.
 KW Hydrolase; Serine protease; Transit peptide: Chloroplast; Thylakoid.
 FT TRANSIT 1 ?
 FT CHAIN 104 437
 FT DOMAIN 152 321
 FT DOMAIN 321 421
 FT ACT_SITE 171 171
 FT ACT_SITE 201 201
 FT ACT_SITE 280 280
 FT CONFLICT 12 23
 FT CONFLICT 36 36
 FT CONFLICT 54 54
 FT CONFLICT 60 60
 FT CONFLICT 64 64
 FT CONFLICT 68 69
 FT CONFLICT 355 355
 FT CONFLICT 381 381
 FT CONFLICT 416 416
 SQ SEQUENCE 437 AA; 46213 MW; 1497B1AB3F5FF2A4 CRC64;
 Query Match 2.3%; Score 9; DB 1; Length 437;
 Best Local Similarity 100.0%; Pred. No. 0.31;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 87 VFAIGNPFG 95
 DB 231 VFAIGNPFG 239
 RESULT 12
 HTRA_BACSU STANDARD; PRT; 449 AA.
 AC O34358;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like htra (EC 3.4.21.-).
 GN HTRA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Devine K.M.;
 RT "Sequence of the Bacillus subtilis genome between xlyA and ykor.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DDbJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brulet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Oglwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).
 RN [3]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX PubMed=10692364;
 RA Noone D., Howell A., Devine K.M.;
 RT "Expression of ykda, encoding a Bacillus subtilis homologue of HtrA,
 RT is heat shock inducible and negatively autoregulated.";
 RL J. Bacteriol. 182:1592-1599(2000).
 RN [4]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX PubMed=11133960;
 RA Noone D., Howell A., Collety R., Devine K.M.;
 RT "Ykda and yvta, HtrA-like serine proteases in Bacillus subtilis,
 RT engage in negative autoregulation and reciprocal cross-regulation of
 RT ykda and yvta gene expression.";
 RL J. Bacteriol. 183:654-663(2001).
 RN [5]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX PubMed=11555295;
 RA Hyrylaenen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
 RA Vitikainen M., Sarvas M., Pragai Z., Bron S., van Dijl J.M.,
 RA Kontinen V.P.;
 RT "A novel two-component regulatory system in Bacillus subtilis for the
 RT survival of severe secretion stress.";
 RL Mol. Microbiol. 41:1159-1172(2001).
 CC -!- FUNCTION: May be involved in processing, maturation, or secretion
 CC of extracellular enzymes.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -!- INDUCTION: Transcription is cspS dependent. Induced by heat shock
 CC during exponential growth and by heterologous amyloses at the
 CC transition phase of the growth cycle. Negatively regulates its own
 CC expression during exponential growth and during heat shock.
 CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
 CC of yvta, especially during stress conditions.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: AJ002571; CAA05570.1; -;
 DR EMBL: Z99110; CAB13147.1; -;
 DR MEROPS: S01.273; -;
 DR Subtilist: BG12608; htra.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50106; PDZ; 1.

KW Hydrolase; Protease; Serine protease; Heat shock; Transmembrane;
KW Complete proteome.
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 67 POTENTIAL.
FT DOMAIN 68 449 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 103 108 POLY-SER.
FT DOMAIN 146 152 POLY-SER.
FT DOMAIN 348 437 PDZ.
FT ACT_SITE 179 179 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 209 209 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 449 AA; 47713 MW; E12B07A9018EE414 CRC64;

Query Match 2.3%; Score 9; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 0.32; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 GIGFAIPSN 168
Db 316 GIGFAIPSN 324
|||||

RESULT 13
DEGP_BUCAI
ID DEGP_BUCAI STANDARD; PRT; 478 AA.
AC P57322;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR BU228.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS".
RL Nature 407:81-86(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001118; BAB12943.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 478 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 116 254 CATALYTIC.
FT DOMAIN 281 372 PDZ 1.
FT DOMAIN 387 469 PDZ 2.
FT ACT_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).

SQ SEQUENCE 478 AA; 52230 MW; 868E8732CAC50629 CRC64;

Query Match 2.3%; Score 9; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 0.34; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 RALGSGVII 17
Db 113 RALGSGVII 121
|||||

RESULT 14
DEGP_BUCAP
ID DEGP_BUCAP STANDARD; PRT; 478 AA.
AC O85291;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98353428; PubMed=9688822;
RA Thao M.L., Baumann P.;
RT "Sequence analysis of a DNA fragment from Buchnera aphidicola (Aphid endosymbiont) containing the genes dapD-htra-llvi-llvh-ftsl-ftsi-murE".
RL Curr. Microbiol. 37:214-216(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC
CC EMBL; AF060492; AAC32331.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolase; Serine protease; Repeat; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 478 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 116 254 CATALYTIC.
FT DOMAIN 281 372 PDZ 1.
FT DOMAIN 387 469 PDZ 2.
FT ACT_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 478 AA; 51303 MW; C044824F7EF4E98E CRC64;

Query Match 2.3%; Score 9; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 0.34; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 YENFIQTDA 126
Db 223 YENFIQTDA 231
|||||

RESULT 15

```
DEGP_RICCN
ID DEGP_RICCN STANDARD; PRT; 508 AA.
AC Q92JAI;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR RC0186.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RA MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE008583; RAL02704.1; ALT_INIT.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00889; trypsin; 1.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 508
FT DOMAIN 119 284
FT DOMAIN 286 377
FT DOMAIN 413 497
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 164 164 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 508 AA; 55599 MW; D2F53A690ECD0AD7 CRC64;

Query Match 2.3%; Score 9; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 GIGFAIPSN 168
Db 269 GIGFAIPSN 277
|||||||
```

Search completed: November 18, 2002, 11:35:28
Job time : 14 secs